

Université
de Toulouse

THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 2 Le Mirail (UT2 Le Mirail)

Discipline ou spécialité :

Sciences du langage

Présentée et soutenue par :

Marianne Vergez-Couret

le : 15 novembre 2010

Titre :

Étude en corpus des réalisations linguistiques de la relation d'Élaboration

Ecole doctorale :

Comportement, Langage, Education, Socialisation, COgnition (CLESCO)

Unité de recherche :

CLLE-ERSS (UMR 5263)

Directeur(s) de Thèse :

Myriam Bras - Professeur, Université de Toulouse 2 Le Mirail

Marie-Paule Péry-Woodley - Professeur, Université de Toulouse 2 Le Mirail

Rapporteurs :

Cathrine Fabricius-Hansen - Professeur, Université d'Oslo

Catherine Schnedecker - Professeur, Université de Strasbourg

Autre(s) membre(s) du jury

Nicholas Asher - Directeur de recherche CNRS, Université de Toulouse 3 Paul Sabatier

Liesbeth Degand - Professeur, Université de Louvain-La-Neuve

Étude en corpus des réalisations linguistiques de la relation
d'*Élaboration*

Marianne Vergez-Couret

15 novembre 2010

Remerciements

Je tiens à remercier un grand nombre de personnes qui m'ont soutenue durant ces quatre années de thèse.

Tout d'abord, je remercie, avec toute mon amitié, mes deux directrices de thèse, Marie-Paule Péry-Woodley et Myriam Bras pour le suivi complémentaire qu'elles m'ont apporté depuis mon Master, pour l'aide qu'elles m'ont apportée en toutes circonstances (je pense notamment aux réunions téléphoniques Bretagne/Toulouse et Strasbourg/Toulouse sous la pluie), pour leur écoute et leurs conseils.

J'adresse mes plus chaleureux remerciements à Cathrine Fabricius-Hansen et Catherine Schnedecker qui ont accepté d'être les rapporteurs de cette thèse. Je les remercie d'avoir évalué mon travail et de m'avoir apporté des remarques et critiques qui me permettront d'avancer dans ma réflexion.

Je remercie Nicholas Asher qui a merveilleusement endossé le rôle de Président du Jury le jour de la soutenance. Je le remercie également pour avoir été présent tout le long de la thèse. Je remercie enfin Liesbeth Degand et souligne la patience dont elle a fait preuve lors des essais techniques pour la vidéo-conférence.

Le laboratoire CLLE-ERSS dans lequel j'ai eu la chance de travailler pendant le doctorat est un vivier de personnes exceptionnelles. D'abord, je tiens à remercier ceux qui ont, tour à tour, été les directeurs de cette équipe et qui ont toujours été soucieux du bien-être des doctorants : Jacques Durand, Anne Condamines, Jean-François Bonnefon et Fabio Montermini. À l'ERSS, j'ai été entourée d'enseignants et de chercheurs enthousiastes, toujours de bonne humeur, empathiques et à l'écoute. Je remercie plus particulièrement : Anne Le Draoulec (pour sa présence depuis le Master), Josette Rebeyrolle et Franck Sajous (qui ont été toujours présents et notamment pendant les crises et paniques de fin de thèse), Cécile Fabre, Ludovic Tanguy (qui m'a montré la voie de la recherche), Nabil Hathout, Andrée Borillo, Michel Aurnague, Claudine Garcia-Debanc. La bonne humeur du laboratoire doit également beaucoup à Laurence Lamy et Nathalie Moulic !

Merci aux collègues de l'IRIT et particulièrement un grand merci à Laure Vieu (pour ces commentaires constructifs du chapitre 3) et à Philippe Muller (pour tous les échanges mails, facebook et les réunions ANNODIS).

Ces quatre années de thèse m'ont également permis de rencontrer de véritables amis. Je dois les remercier pour leur gentillesse, leur générosité et leur joie de vivre : Mai Ho-Dac, la grande sœur qui me guida vers les petits chemins tortueux (et je l'en remercie), Laurent Prévot (le grand frère), Clémentine Adam (avec qui ce fut un plaisir de travailler – on remet ça quand tu veux!!), François Morlane-Hondère (avec qui ce fut un plaisir de partager le bureau), Buddy Dirat (pour le travail accompli en tant que représentant des doctorants et surtout pour l'ami qu'il est devenu), Aurélie Guerrero (et toutes les heures skype passées

ensemble les soirs de rédaction strasbourgeois), Christelle Pêcher (et merci le CIES, il fallait bien que ça serve à quelque chose!), Aurélie Picton (avec qui j'espère toujours partager un bureau, même dans longtemps), Christel Le Bellec (toujours souriante), Lionel Fontan (formidable de gentillesse et de générosité de cœur), Charlotte Alazard (pour toutes ses psychanalyses de fin de thèse), Marjorie Raufast (qui ajoute du fun là où elle passe), Stéphanie Lopez, Nathalie Dehaut (pour son «je crois en moi, je crois en mon pouvoir, je porte en moi le triomphe!»), Marion Laignelet, Christophe Pimm, Vladimir Moscovici, Fanny Lalleman, Marie-France Roquelaure, Caitlin V. Smith, Daniella Muller.

Enfin, une mention toute spéciale à Angelina Aleksandrova (si si), rencontrée en toute fin de thèse mais qui est devenue très rapidement une amie et qui m'a soutenue dans les derniers moments qui n'ont pas été toujours faciles !

Un grand merci à Nadou qui a veillé à ce que je sorte le nez de mes bouquins pour courir sur les tapis de gym !

Je tiens enfin à remercier toute ma famille pour leur soutien et leur présence infaillible. Maman et Denis qui ont écouté et cuisiné durant mes dernières semaines à Toulouse. Papa et sa théorie de la montagne. Mes sœurs, Line et Laurine, mes taties, tontons, cousins et cousines. Papi et Mamie. Le clan des M : Michelle, Marc, Magali, Matthieu et Marianne.

Fred (avec tout mon amour).

Un grand MERCI à vous tous !

Je crois qu'il profita, pour son évasion, d'une migration d'oiseaux sauvages. Au matin du départ il mit sa planète bien en ordre. Il ramona soigneusement ses volcans en activité. Il possédait deux volcans en activité. Et c'était bien commode pour faire chauffer le petit déjeuner du matin. Il possédait aussi un volcan éteint. Mais, comme il disait, «On ne sait jamais!» Il ramona donc également le volcan éteint. S'ils sont bien ramonés, les volcans brûlent doucement et régulièrement, sans éruptions. Les éruptions volcaniques sont comme des feux de cheminée. Évidemment sur notre terre nous sommes beaucoup trop petits pour ramoner nos volcans. C'est pourquoi ils nous causent des tas d'ennuis. Le petit prince arracha aussi, avec un peu de mélancolie, les dernières pousses de baobabs. Il croyait ne jamais devoir revenir. Mais tous ces travaux familiers lui parurent, ce matin-là, extrêmement doux. Et, quand il arrosa une dernière fois la fleur, et se prépara à la mettre à l'abri sous son globe, il se découvrit l'envie de pleurer.

(*Le Petit Prince*, Chapitre IX, Antoine de Saint-Exupéry, 1943)

Sommaire

Introduction	xv
I La relation d'Élaboration au cœur des structures du discours	1
1 La structure du discours	5
1.1 Cohérence du discours	5
1.1.1 Production <i>vs.</i> réception des textes	6
1.1.2 Principes de la cohérence discursive	7
1.1.3 Cohésion	8
1.1.4 Deux sources implicites de cohérence	8
1.1.5 Bilan	10
1.2 Structures discursives et relations de discours	10
1.2.1 Nature des segments de discours	11
1.2.2 Nombre des relations de discours	11
1.2.3 Nature des relations de discours	12
1.2.4 Structure hiérarchique	13
1.2.5 Reconnaissance des relations de discours	14
1.2.6 Positionnement	14
1.3 La SDRT	15
1.3.1 Représenter la sémantique du discours en DRT	15
1.3.2 Représenter la structure du discours en SDRT	17
1.4 Structures discursives et applications de TAL	28
1.4.1 Distance entre approches théoriques et applications de TAL	28
1.4.2 Travaux récents sur la détection automatique des relations de discours	29
1.5 Bilan et positionnement	30
2 Les descriptions de l'élaboration	33
2.1 Historique de la notion d'élaboration	34
2.1.1 Grammaires traditionnelles	34
2.1.2 Prédicats rhétoriques	36
2.1.3 Analyse conversationnelle	39
2.1.4 Bilan	41
2.2 État de l'art des définitions données à l'Élaboration	41
2.2.1 Points communs et divergences	42

2.2.2	Approches fonctionnelles	44
2.2.3	Approches formelles	57
2.3	<i>Élaboration</i> en génération automatique de texte	65
2.4	Synthèse	67
3	La relation d'<i>Élaboration</i> en SDRT	71
3.1	Définitions	72
3.1.1	Définitions (Asher, 1993)	72
3.1.2	Calcul des relations temporelles (Lascarides et Asher, 1993)	78
3.1.3	Descriptions (Asher et Lascarides, 2003)	79
3.2	Inférer <i>Élaboration</i>	85
3.2.1	Le prédicat <i>Subtype_D</i>	85
3.2.2	Les cadres de discours	87
3.3	Une <i>Élaboration</i> , des <i>Élaborations</i> ?	89
3.3.1	Distinguer <i>Élaboration d'Entité</i> et <i>Élaboration d'Éventualité</i>	89
3.3.2	Distinguer <i>Élaboration d'état</i> et <i>Élaboration d'événement</i>	92
3.3.3	Distinguer <i>Particularisation</i> et <i>Généralisation</i>	93
3.4	Synthèse	95
4	La signalisation des relations de discours	99
4.1	Marqueurs du discours dans les théories du discours	100
4.1.1	Caractère labile des marqueurs du discours	100
4.1.2	Rôle des marqueurs du discours dans les théories du discours	104
4.1.3	Synthèse sur indices et marqueurs du discours	111
4.2	Autres indices de la signalisation discursive	112
4.2.1	Des indices moins étudiés	112
4.2.2	Corrélation entre indices et relations de discours	112
4.3	Rôle des indices de la signalisation discursive en TAL	113
4.3.1	La génération automatique de texte	114
4.3.2	Le résumé automatique de texte	115
4.4	La signalisation de la relation d' <i>Élaboration</i>	116
4.4.1	Une relation non marquée et une relation par défaut	116
4.4.2	Des études sur la signalisation de la relation d' <i>Élaboration</i>	118
4.5	Bilan et positionnement	122
II	Analyse de marqueurs potentiels de l'<i>Élaboration</i> en SDRT	125
5	Les adverbess paradigmatissants	129
5.1	Les adverbess paradigmatissants	130
5.1.1	Définitions	130
5.1.2	Position de l'adverbe et nature du paradigme	130
5.1.3	Une classe d'adverbess à part	132
5.1.4	Rôle de focalisation <i>vs.</i> rôle paradigmatissant	134
5.2	Rôle discursif des adverbess en position intraphrastique	138
5.2.1	<i>Notamment</i>	138

5.2.2	<i>Particulièrement</i>	146
5.2.3	<i>En particulier</i>	147
5.2.4	<i>Précisément</i>	148
5.2.5	Conclusion sur les emplois intraphrastiques	149
5.3	Rôle discursif des adverbes en position initiale détachée	150
5.3.1	Spécificité de la position initiale détachée	150
5.3.2	<i>Notamment, en particulier</i>	152
5.3.3	<i>Plus particulièrement</i>	154
5.3.4	<i>Plus précisément</i>	162
5.3.5	Conclusion sur les emplois à l'initiale	165
5.4	Conclusion	166
6	Des marqueurs structurels aux structures énumératives	169
6.1	Marqueurs d'ouverture	170
6.1.1	Double attachement	171
6.1.2	Attachement au contexte gauche	174
6.1.3	Synthèse	179
6.2	Structures énumératives	181
6.2.1	Les structures énumératives dans les théories du discours	181
6.2.2	Représentation en SDRT des structures énumératives	182
6.2.3	Synthèse	193
6.3	Structures énumératives à deux temps	194
6.3.1	Description des structures énumératives à deux temps	194
6.3.2	Des structures similaires dans les théories du discours	196
6.3.3	Traitement des structures énumératives à deux temps	199
6.3.4	Synthèse	205
6.4	Conclusion	205
7	Une structure syntaxique, le gérondif	207
7.1	La construction gérondive	208
7.1.1	Aspects syntaxiques	208
7.1.2	Aspects sémantiques	210
7.2	La typologie de Halmøy (1982)	211
7.2.1	Repère temporel (RT)	211
7.2.2	Relation de cause, de condition ou de moyen (A)	211
7.2.3	Relation d'inclusion ou d'équivalence (A')	213
7.2.4	Relation de concomitance (B)	214
7.2.5	Relation d'hyponymie (B')	216
7.3	Représentation discursive du gérondif	218
7.3.1	Configuration A <résultat, moyen>	218
7.3.2	Configuration A'	223
7.3.3	Configuration B	224
7.3.4	Configuration B'	228
7.4	Conclusion	229

III	Approche en corpus de l'<i>Élaboration</i>	231
8	Présentation du corpus ANNODIS	235
8.1	Préparation de l'annotation	236
8.1.1	Objets à annoter	236
8.1.2	Annotation exploratoire	239
8.2	Déroulement de l'annotation et manuels	239
8.2.1	Perspective descendante	239
8.2.2	Perspective ascendante	240
8.2.3	Outil d'annotation	252
8.2.4	Exploitation des données	252
8.3	Perspectives	252
9	<i>Élaboration</i> au cœur du corpus ANNODIS	255
9.1	Analyse quantitative	256
9.1.1	Relations les plus fréquentes	256
9.1.2	Accord entre annotateurs sur les relations d' <i>élaboration</i>	256
9.1.3	Divergences	258
9.2	Analyse qualitative	259
9.2.1	<i>Via</i> les marqueurs de la partie 2	259
9.2.2	<i>Via</i> l'annotation des structures énumératives	275
9.2.3	<i>Via</i> l'annotation standard	285
9.2.4	Bilan	299
9.3	Conclusion	300
10	Identification automatique d'<i>Élaboration</i> en corpus	303
10.1	De la cohésion lexicale à la cohérence discursive	304
10.1.1	Répétition lexicale et chaînes lexicales	304
10.1.2	Indices lexicaux pour l'identification d' <i>Élaboration</i>	305
10.1.3	Choix de la ressource lexicale	307
10.2	Une expérience de combinaison	309
10.2.1	Combinaison d'indices	309
10.2.2	Principes de l'expérience et validation expérimentale	310
10.2.3	Bilan de l'expérimentation	312
10.3	Conclusions et perspectives	313
	Conclusion	315

Need I elaborate ?

Liste d'abréviations

ANNODIS	ANNOtation DIScursive
CMLF	Congrès mondial de linguistique française
CNRTL	Centre national de ressources textuelles et lexicales
DISCOR	DIScourse Structure and COReference Resolution
DRS	Discourse Representation Structure (Structure des représentations discursives)
DRT	Discourse Representation Theory (Théorie des représentations discursives)
DS	Discourse segment (Segment de discours)
ER	Est Républicain
GSV	Gérondif sans voisin
ILF	Institut de langue française
INT	Introduction de nouveau topique
JFLS	Journal of French Language Studies
LEXCONN	French LEXicon of Discourse CONNectives
LPTS	Linguistic and psycholinguistic approaches to text structuring (Approche linguistique et psycholinguistique de la structure des textes)
MAT	Modèle d'architecture textuelle
MDC	Maximize Discourse Coherence (Maximiser la cohérence du discours)
MAI	Marqueur d'autre item
MPI	Marqueur de premier item
OMC	Oslo multilingual corpus
PID	Position initiale détachée
RST	Rhetorical Structure Theory (Théorie des structures rhétoriques)
SDRS	Segmented Discourse Representation Structure (Structure des représentations discursives segmentées)
SDRT	Segmented Discourse Representation Theory (Théorie des représentations discursives segmentées)
SLDM	Second-level Discourse Markers (Marqueurs du discours de second niveau)
SE	Structure énumérative
SG	Syntagme gérondif
SN	Syntagme nominal
SP	Syntagme prépositionnel
SPRIK	Språk in contrast (Langues en contraste)
SUR	Segment à unité référentielle
SV	Syntagme verbal

TAL Traitement automatique des langues

UDE Unité de discours élémentaire

VR Verbe régissant

VV Verbes voisins

VVOV Verbes voisins et objets voisins

Wiki Wikipedia

Introduction

Nous nous interrogeons dans cette thèse sur les réalisations linguistiques des relations de discours en considérant la relation de discours *Élaboration* que nous illustrons avec l'épigraphe de notre thèse :

Au matin du départ il mit sa planète bien en ordre. Il ramona soigneusement ses volcans en activité. (...) Il ramona donc également le volcan éteint. (...) Le petit prince arracha aussi, avec un peu de mélancolie, les dernières pousses de baobabs.

La première phrase décrit un événement «mettre sa planète bien en ordre». Les événements «ramoner ses volcans en activité», «ramoner le volcan éteint» et «arracher les dernières pousses de baobabs», introduits par les phrases suivantes, apportent des détails sur celui-ci en décrivant des événements participant à la mise en ordre de la planète.

Entre la première et la deuxième phrase, nous notons l'absence d'un quelconque marqueur lexical signalant la relation. Nous pouvons contraster cet exemple avec une phrase contenant le marqueur *pour commencer* qui explicite la relation :

Au matin du départ il mit sa planète bien en ordre. Pour commencer, il ramona soigneusement ses volcans en activité.

En l'absence de marques lexicales, d'autres éléments permettent l'interprétation de l'*Élaboration*. C'est sur la base de connaissances du monde et de connaissances construites à la lecture du texte des trente premières pages du livre que l'on comprend les événements de «ramoner ses volcans» et d'«arracher les dernières pousses de baobabs» comme élaborant celui de «mettre sa planète bien en ordre» en se basant sur les analogies entre «mettre en ordre une planète et une maison», «ramoner des volcans et des cheminées» et «arracher les dernières pousses de baobabs et les mauvaises herbes». La poursuite de l'élaboration est, par ailleurs, facilitée par les adverbes d'addition *également* et *aussi*. Nous avons ici une élaboration que l'on pourrait qualifier de *poétique*. Les exemples considérés dans cette thèse, issus d'articles de presse, d'articles scientifiques et d'articles de Wikipedia, seront bien moins poétiques mais mettront tout aussi bien en évidence l'importance des connaissances du monde et du contexte discursif pour l'interprétation des relations de discours.

Pour observer les réalisations linguistiques de la relation d'*Élaboration*, nous allons faire appel aux approches et modèles de la structure du discours à des fins de description linguistique, de modélisation linguistique, d'exploration sur corpus et de traitement informatique de cette structure. Notre approche, qui multiplie les points de vue, s'inscrit dans une tendance plus générale d'échange et de confrontation interdisciplinaire des travaux sur la

structure du discours, comme en témoigne la création récente de la revue *Discours*¹ et d'un atelier spécialisé biannuel *Multidisciplinary Approaches to Discourse* (MAD).

Nos travaux ont eu pour cadre le laboratoire CLLE-ERSS (Cognition, Langues, Langage et Ergonomie - Équipe de Recherche en Syntaxe et Sémantique). Ils s'inscrivent dans deux thématiques : «Structures discursives» de l'axe S'caladis (Structures Sémantiques : des catégories lexicales au discours) et «Discours et Documents» de l'axe TAL (Traitement Automatique des Langues). Ils ont ainsi pu bénéficier de l'expérience des membres de ces deux thématiques concernant la description et la modélisation linguistique d'une part et les traitements sur corpus d'autre part.

Nous avons rejoint en 2006, lors de notre année de Master 2, le projet ILF «Relations de cohérence et fonctionnement des anaphores» (2003-2006), coordonné par François Cornish et réunissant les laboratoires CLLE-ERSS de l'université de Toulouse 2 Le Mirail et EA 1339 de l'université Marc Bloch de Strasbourg. Un des objectifs de ce projet était d'examiner les interactions existant entre le choix d'une relation de discours et le fonctionnement de marqueurs lexicaux et d'expressions anaphoriques. Le projet s'est focalisé sur la relation d'*Élaboration*, jusqu'alors peu étudiée. Notre rôle, quand nous avons été intégrée dans le projet, était de compléter la liste des marqueurs potentiels de cette relation (Vergez-Couret, 2006).

Notre équipe de recherche s'est ensuite impliquée dans le projet ANNODIS² (2007-2010). Ce projet, coordonné par Marie-Paule Péry-Woodley, réunit trois laboratoires de recherche : CLLE-ERSS, Toulouse ; IRIT, Toulouse et GREYC, Caen. Nous y avons pris une part active en particulier avec nos travaux sur l'*Élaboration*. Son objectif final est de mettre à disposition de la communauté scientifique un corpus enrichi d'annotations discursives et d'outils informatiques permettant l'annotation et l'exploration d'un tel corpus.

Le choix de la relation d'*Élaboration*, dans le cadre du projet ILF, s'explique à l'origine par la faible quantité d'études concernant la relation. Ce constat peut s'expliquer par le fait que cette relation n'a pas de marqueur prototypique, contrairement, par exemple, à la relation d'*Explication* qui a pour marqueur prototypique *parce que* et la relation de *Résultat* qui a pour marqueur prototypique *donc*. Pourtant, cette relation est très présente dans les textes comme le montreront nos données. Et, elle occupe une place centrale dans les structures du discours pour des raisons que nous mettrons en évidence dans cette thèse.

En nous focalisant sur les réalisations linguistiques de la relation d'*Élaboration*, nous poursuivons un double objectif. Nous souhaitons proposer des descriptions linguistiques de marqueurs, d'indices et de configurations d'indices pouvant enrichir un modèle du discours et permettre une identification automatique ou semi-automatique de la relation. Les premiers enjeux sont essentiellement empiriques. Nous souhaitons faire progresser des analyses descriptives sur cette relation. Les seconds enjeux sont plus applicatifs. Repérer automatiquement les relations de discours constitue un défi considérable pour des applications de

¹Discours, Revue de linguistique, psycholinguistique et informatique (<http://discours.revues.org/>).

²Projet financé par l'ANR - Programme Sciences Humaines et Sociales Appel 2007 «Corpus et outils de la recherche en sciences humaines et sociales».

traitement automatique des langues (ci-après TAL) telles que la recherche d'information, la génération automatique de texte, le résumé automatique de texte...

Pour poursuivre cet objectif, nous adoptons une approche pragmatico-sémantique de la structure discursive en prenant pour modèle la Théorie des Représentations Discursives Segmentées (*Segmented Discourse Representation Theory*, ci-après SDRT). Ce cadre vise la représentation du contenu sémantique des propositions, de la segmentation et de l'organisation de la macro-structure du discours *via* des relations de discours.

L'absence de marqueur type pour la relation d'*Élaboration* fait de son étude un champ d'observation propice à la confrontation des deux approches onomasiologique et sémasiologique. La première méthode, dite onomasiologique, part de la relation de discours pour envisager ses réalisations linguistiques tandis que la seconde méthode, dite sémasiologique, part d'un marqueur potentiel pour en déterminer son rôle dans la mise en œuvre de la relation de discours.

Notre point de départ, pour cette étude de l'*Élaboration*, est d'envisager en largeur les réalisations linguistiques possibles de la relation. Nous qualifions cette première étape d'approche onomasiologique introspective. Il s'agit de déterminer, par l'introspection ou par l'exploration sporadique de textes, quels sont les éléments qui participent à l'interprétation de la relation d'*Élaboration*. Une fois de tels éléments mis au jour, l'approche sémasiologique permet d'appréhender leurs fonctionnements dans la mise en œuvre de la relation.

La mise en place de l'analyse sémasiologique nécessite l'utilisation d'un corpus «nu» que l'on interroge au moyen d'un concordancier. Ces méthodes font intervenir un dispositif léger, satisfaisant pour travailler sur une ou plusieurs formes lexicales ou syntaxiques, permettant d'accéder à des occurrences des marqueurs potentiels pour déterminer s'ils sont en corrélation permanente ou occasionnelle avec la relation de discours étudiée. Il est ainsi possible de déterminer des contraintes pesant sur l'interprétation du marqueur. Le recours à des manipulations linguistiques permet de tester plus finement le rôle que celui-ci (dans son contexte) joue dans la mise en œuvre de la relation.

La SDRT accorde une place majeure aux processus interprétatifs, impliquant de porter une attention très fine aux réalisations linguistiques à travers les marqueurs lexicaux, les structures syntaxiques, les marqueurs aspectuo-temporels, la sémantique lexicale, les connaissances du monde et leurs interactions. Cette précision a entraîné, pour la constitution du modèle (Asher, 1993; Asher et Lascarides, 2003), à la fois la limitation des données analysées et la simplification des exemples construits (permettant de mettre en avant une propriété décisive dans la mise en œuvre d'une relation de discours).

L'originalité de notre travail est de proposer la confrontation de la théorie à des données attestées en s'appuyant sur des descriptions les plus fines possibles. Ce travail peut être mené à bien parce que nous ne regardons qu'une seule relation de discours, la relation d'*Élaboration*. Un examen attentif des descriptions de la relation dans plusieurs travaux et plusieurs théories a permis de dégager une multitude de configurations sémantiques sous l'étendard de l'*Élaboration*. Nous verrons que l'analyse de données attestées va permettre d'enrichir le modèle en fournissant de nouvelles règles d'inférence pour la relation.

Dans cette optique, nos travaux se situent à l'interface d'une étude empirique et d'une étude théorique. Partant d'un *marqueur potentiel*, des descriptions sémantiques fines per-

mettent de déterminer son rôle dans la mise en œuvre d'une relation de discours. Nous réservons le terme de *marqueur de relations de discours*, dans une perspective d'analyse, aux éléments dont la ou les fonctions discursives ont été établies. Les *marqueurs d'une relation de discours* sont, quant à eux, ceux que l'on retrouve dans les règles d'inférence des relations de discours de la SDRT.

Nous proposons des descriptions linguistiques qui donnent lieu dans certains cas à la formalisation, en proposant de nouvelles règles d'inférence pouvant enrichir directement le modèle de la SDRT. Mais plus globalement, toutes nos analyses descriptives, même non formalisées, ont pour visée de compléter et d'enrichir le modèle.

Pour systématiser les apports de l'approche onomasiologique, il est nécessaire d'avoir accès à un corpus annoté en relations de discours. Nous avons participé à la construction d'un tel corpus dans le cadre du projet ANNODIS, notamment en contribuant à l'élaboration des manuels de segmentation et d'annotation. Le corpus ANNODIS est le premier corpus pour le français fournissant ce type d'annotation. Cette approche permet d'aborder les réalisations de la relation sans porter de jugement sur la façon dont celle-ci pourrait être réalisée. Adopter cette démarche est quasiment rendu nécessaire par les manifestations de la relation d'*Élaboration* qui ne passent que très rarement par la mention d'un marqueur lexical. Toutefois, interroger un tel corpus est moins trivial qu'interroger un corpus «nu» à l'aide d'un concordancier. Nous mettrons donc en place une méthodologie pour l'exploration de ce corpus.

Nous verrons comment cette approche nous a permis d'accéder, *via* la relation de discours, à des indices. Dans un cadre idéal, nous pouvons imaginer que la mise en corrélation d'indices et de la relation de discours peut se suffire à elle-même. Mais en l'état actuel du corpus, seule une analyse sémasiologique peut permettre de découvrir si ces indices sont des marqueurs, ou pour être plus exact dans quels contextes ces indices sont des marqueurs de la relation. Nous verrons en quoi ces deux types d'approches – sémasiologique et onomasiologique – peuvent être complémentaires pour l'étude d'une relation de discours.

Dans une perspective d'identification automatique ou semi-automatique de la relation, marqueurs et indices constituent des sources d'informations très riches, facilement accessibles automatiquement. Dès lors que nous cherchons à repérer automatiquement tout ou partie de la structure du discours, l'utilisation d'une ressource mettant en corrélation marqueurs du discours et relations du discours semble inévitable. Mais la construction d'une telle ressource se heurte à la complexité du rôle et de la place des marqueurs dans l'interprétation. Nous le savons bien, les marqueurs du discours sont ambigus et leur désambiguïsation passe généralement par la prise en compte de leur contexte. Nous souhaitons mettre au point une méthode qui permette de mettre au jour des faisceaux d'indices, repérables automatiquement, de la relation d'*Élaboration*.

★★★

La thèse est organisée en trois parties. La première partie est consacrée à l'état de l'art. Nous proposons, dans le premier chapitre, de présenter les théories et relations du discours. Nous y présentons les principaux fondements de la SDRT. Un examen attentif

des descriptions de la relation d'*Élaboration* dans plusieurs travaux et théories est présenté dans le chapitre 2. Puis nous revenons au cadre de la SDRT pour présenter les descriptions de la relation d'*Élaboration* qui seront à la base de nos analyses. Pour terminer l'état de l'art, nous allons poser la problématique du caractère labile des marqueurs du discours et de la signalisation des relations de discours en général (chapitre 4).

La seconde partie est consacrée à l'analyse de marqueurs potentiels de l'*Élaboration*. Nous ne fournirons pas une liste exhaustive de ces derniers. Notre objectif est, ici, de fournir les analyses les plus approfondies possibles du rôle des marqueurs considérés dans la mise en œuvre de la relation, nous limitant, pour des raisons de place et de temps évidentes, à une liste fermée. Dans le chapitre 5, nous analyserons quatre adverbes paradigmatiques *notamment*, *en particulier*, *particulièrement* et *précisément*, leur rôle étant, à première vue, d'apporter des éléments de détails. Pour étudier leur rôle dans la mise en œuvre de la relation d'*Élaboration*, nous étudierons deux configurations précises dans lesquelles les marqueurs portent sur des éventualités (en position intraphrastique ou en position initiale). Dans le chapitre 6, nous analyserons les structures énumératives. Celles-ci sont repérables *via* une configuration d'indices (amorces, marqueurs d'item...) et forment des «blocs» assumant une fonction rhétorique particulière. L'objectif est de découvrir si ces structures sont isomorphes à des structures élaboratives. Enfin dans le chapitre 7, nous étudierons une structure syntaxique, le gérondif. La diversité interprétative de cette construction justifie une analyse sémantique. Nous prolongerons l'analyse de Halmøy (2003) en proposant une analyse en relations de discours. Nous montrerons la complexité du rôle des marqueurs dans l'interprétation des discours et la mise en œuvre des relations de discours en proposant des analyses détaillées sur de nombreux exemples attestés pour chaque marqueur.

La troisième partie concerne l'analyse de la relation d'*Élaboration* en adoptant le point de vue de la linguistique de corpus. Le chapitre 8 est consacré à la présentation du corpus ANNODIS. Puis, le chapitre 9 présente notre exploration préliminaire des données du corpus. Nous y présentons l'approche onomasiologique de recherches d'indices de la relation d'*Élaboration*. Ainsi, nous cherchons à présenter la meilleure méthodologie tirant au mieux parti des deux approches sémasiologique et onomasiologique. Enfin, notre dernière contribution dans la thèse vise la détection automatique d'occurrences de la relation d'*Élaboration* en proposant une méthode de combinaison de plusieurs indices (chapitre 10).

Première partie

La relation d'*Élaboration* au cœur des structures du discours

Cette partie est composée de quatre chapitres. Nous proposons dans le premier chapitre un état de l'art des théories du discours. Nous présentons plus en détail le cadre théorique de la SDRT et les raisons pour lesquelles nous l'avons choisi pour nos analyses. Les deux chapitres suivants sont consacrés à la description de la relation d'*Élaboration*. D'abord, nous proposons une présentation de la notion d'élaboration en partant de travaux anciens jusqu'aux définitions de la relation données dans les théories du discours (chapitre 2). Puis, nous présentons en détail les évolutions de la relation d'*Élaboration* dans le cadre de la SDRT et les enjeux actuels de l'étude de cette relation (chapitre 3). Enfin, nous présentons des études qui ont porté sur la signalisation des relations de discours avec pour question : «comment les relations de discours sont-elles signalées dans les textes» (chapitre 4). Nous nous focaliserons d'abord sur le rôle et la place des marqueurs du discours et des autres indices de la signalisation discursive dans les théories du discours ; puis plus spécifiquement sur la signalisation de la relation d'*Élaboration*.

Autour de la relation d'*Élaboration*, nous nous interrogeons sur la façon dont les relations de discours sont exprimées dans les textes. Nous souhaitons rechercher des marqueurs et des indices de la relation avec un double objectif. D'une part, nous souhaitons proposer des descriptions linguistiques de marqueurs permettant d'enrichir le modèle de la SDRT. Et d'autre part, nous souhaitons proposer des patrons obtenus par la combinaison d'indices pour l'identification automatique de la relation d'*Élaboration*. Tout au long de cette première partie, et de façon transversale, nous présenterons la place des théories du discours et le rôle des marqueurs et indices du discours dans les applications de traitement automatique des langues (ci-après TAL).

Chapitre 1

La structure du discours

Sommaire

1.1	Cohérence du discours	5
1.1.1	Production <i>vs.</i> réception des textes	6
1.1.2	Principes de la cohérence discursive	7
1.1.3	Cohésion	8
1.1.4	Deux sources implicites de cohérence	8
1.1.5	Bilan	10
1.2	Structures discursives et relations de discours	10
1.2.1	Nature des segments de discours	11
1.2.2	Nombre des relations de discours	11
1.2.3	Nature des relations de discours	12
1.2.4	Structure hiérarchique	13
1.2.5	Reconnaissance des relations de discours	14
1.2.6	Positionnement	14
1.3	La SDRT	15
1.3.1	Représenter la sémantique du discours en DRT	15
1.3.1.1	Une théorie représentationnelle et dynamique	15
1.3.1.2	La représentation du discours	16
1.3.2	Représenter la structure du discours en SDRT	17
1.3.2.1	Unités de discours élémentaires et relations de discours	18
1.3.2.2	Construction des SDRS	20
1.3.2.3	Inférence des relations de discours	23
1.4	Structures discursives et applications de TAL	28
1.4.1	Distance entre approches théoriques et applications de TAL	28
1.4.2	Travaux récents sur la détection automatique des relations de discours	29
1.5	Bilan et positionnement	30

1.1 Cohérence du discours

Un discours n'est pas une simple suite de phrases énoncées. Il existe, d'une part, des marques de *cohésion* linguistique (reprises anaphoriques et lexicales, connecteurs...) qui as-

surent des connexions d'un énoncé à l'autre. Et d'autre part, il existe des liens de *cohérence* entre les énoncés qui font qu'un discours est compris comme un tout. Ce tout n'est pas uniquement construit à partir de la somme des contenus sémantiques de chaque partie et fait intervenir des dépendances, traduites au moyen de la notion de *relation de discours*, entre ces parties. Les théories du discours posent l'hypothèse que toutes les parties d'un discours cohérent sont attachées par (au moins) une relation de discours à une autre partie.

Cette structuration du discours repose sur une multitude de facteurs dont les théories du discours, qui ont pour objectif de modéliser l'organisation structurelle du discours, cherchent à déterminer la nature et les interactions.

1.1.1 Production *vs.* réception des textes

Dans la production ou la réception d'un texte, la construction de cette structure discursive (ou *text representation* chez (Sanders et Spooren, 2001)) est décrite comme une représentation mentale. Lors de la production, cette représentation mentale est à la base de la constitution des segments (choix lexicaux, syntaxiques...), de la hiérarchisation et de la mise en relation de ces derniers. La réalisation linguistique de cette construction mentale impose une linéarisation. Et lors de la réception, à partir de cette réalisation linguistique linéaire, le lecteur construit le sens du discours à partir du contenu sémantique de chaque partie et des dépendances entre ces parties.

Les théories du discours en production décrivent le passage de cette représentation mentale hiérarchisée à un format linéaire imposé par la mise en texte (Fayol, 1997) tandis que les théories du discours en réception décrivent et/ou formalisent des représentations, à partir d'une réalisation linguistique.

Nous avons évoquées jusque là deux notions : le texte (ou réalisation linguistique) et le discours (ou construction mentale), que Cornish (2003) définit ainsi :

I view text (as a non-count noun) as denoting a typical instance of language cum other semiotic devices in use - i.e. occurring in some context and with the intention by the user of achieving some purpose or goal thereby. The term designates the connected sequence of verbal signs and non-verbal signals, vocal as well as non-vocal (i.e. visual, auditory, etc.) signals produced within the context of some utterance act. (...)

Discourse, on the other hand, designates the hierarchically structured, mentally represented sequences of utterance and indexical acts which the participants are engaging in as the communication unfolds. Such sequences have as their raison d'être the accomplishment of some particular overall communicative goal.

(Cornish, 2003, p. 33-34)

Le texte est l'ensemble des *signes verbaux* et des *signes non verbaux* résultant d'un discours produit dans un certain contexte. Les informations linguistiques présentes dans le texte sont de nature très diverse : les signes verbaux sont composés, à l'écrit et à l'oral, des séquences de mots, qui forment des propositions auxquelles sont assignées des informations sémantiques et grammaticales. Les gestes, les mimiques, les rires... constituent à l'oral les signes non verbaux. À l'écrit, les signes non verbaux concernent les propriétés visuelles du texte (Luc, 2000; Péry-Woodley, 2000) : les textes écrits sont matérialisés sur un support

papier ou informatique. Ils sont des objets visuels dont les propriétés jouent également un rôle dans la construction de la cohérence discursive. L'ensemble des signes verbaux et non verbaux fournissent une base sur laquelle reposent les interprétations du lecteur/allocutaire dans l'objectif de produire une représentation de ce texte, i.e. le discours qui est défini comme une structure hiérarchique d'actes de langage produits par un scripteur/locuteur selon certains buts communicatifs dans un contexte donné. C'est à travers ce contexte que le lecteur/allocutaire est capable de connecter les séquences du texte en discours.

1.1.2 Principes de la cohérence discursive

Lors de l'étape de réception du texte, les lecteurs ou allocutaires cherchent à construire une représentation mentale cohérente de ce dernier. Selon Charolles *et al.* (2005), le principe général de cohérence régit les processus interprétatifs :

Imagine you are sitting in the back of a car. A and B are sitting in the front. A says : «The car's making a funny noise». Immediately afterwards, B says : «I haven't got my wallet». Either you think A and B are each following their own train of thought and wording completely independant ideas, or you consider they are engaging in discourse, in which case you need to calculate a relation in order to understand how what B said connects with what A said. For example, B has interpreted A's utterance as indicating that the car will need attending to, with the implication that money will be required.

(Charolles *et al.*, 2005, p. 115)

Lors du traitement d'un nouvel énoncé (résultat d'un acte d'énonciation), le lecteur ou l'allocutaire, considérant que ce qu'on lui dit est cohérent, va produire des inférences pour extraire du texte une interprétation en accord avec sa propre conception d'un discours cohérent (Charolles, 1983).

L'examen d'un texte révèle des expressions indiquant que telle ou telle partie de celui-ci doit être reliée à une autre partie. Charolles (1997) signale que ces expressions guident le lecteur dans la résolution des inférences, ce qui lui permet alors de construire une interprétation :

L'interprétation des discours est soumise à un principe général de cohérence (Charolles 1983, 1994) ou de pertinence (Sperber & Wilson 1986) qui est de nature fondamentalement sémantique et pragmatique. Confronté à une séquence d'énoncés produits à la suite, le destinataire ne peut en effet que chercher à établir des relations entre ces énoncés, vu que, précisément, ils sont énoncés à la suite. L'établissement de ces liens fait appel à des opérations intellectuelles de haut niveau dans laquelle interviennent toutes sortes de compétences linguistiques et non linguistiques. Pour guider l'interlocuteur dans le processus de résolution de problèmes, le locuteur a à sa disposition un vaste ensemble de marques de cohésion qui codent des instructions relationnelles plus ou moins spécifiques.

(Charolles, 1997, p. 3)

1.1.3 Cohésion

L'occurrence de ces expressions relationnelles confère au texte une certaine cohésion. Halliday et Hasan (1976) parlent de *texture*, en référence à la manière dont les fils s'entrecroisent pour former un tissu, pour exprimer l'idée que ces expressions relationnelles sont les fils qui tissent un texte. Selon eux, la cohésion se manifeste au moyen de relations sémantiques non structurelles créant des liens de dépendance entre les unités constitutives du texte. Ces moyens se répartissent en cinq catégories : la *référence* (anaphore et cataphore), la *substitution*, l'*ellipse*, la *conjonction* (coordination et connecteur) et la *cohésion lexicale* (répétition d'un mot, utilisation d'un synonyme, d'un hyperonyme...).

Ces marques de cohésion servent d'instructions indiquant au lecteur la mise en relation d'éléments du texte mais le contenu sémantique de ces marques demeure incomplet et/ou ambigu et leur résolution dépend d'autres aspects discursifs et pragmatiques en lien avec le contexte. A titre d'illustration, nous pouvons citer les anaphores (qui ne pointent pas de façon précise l'antécédent nécessaire à leur résolution) et les connecteurs (qui ne désignent pas avec précision les unités du texte qu'ils relient). Les marques de cohésion fonctionnent comme des *signaux* ou des *déclencheurs* (Cornish, 1990) qui signalent/déclenchent des processus inférentiels pour lesquels les informations du contexte jouent un rôle essentiel.

L'établissement de ces marques de cohésion et de leurs interactions constitue une thématique de recherche encore très fructueuse en analyse linguistique du discours. Mais ces travaux ne peuvent pas, à eux seuls, rendre compte de la cohérence. L'occurrence de marques de cohésion dans un texte n'est ni nécessaire, ni suffisante pour qu'une séquence apparaisse cohérente.

1.1.4 Deux sources implicites de cohérence

D'autres sources permettent de rendre compte de la cohérence d'un discours. Il s'agit de divers types de structures sous-jacentes attribuées au discours et qui sont exploitables pour contribuer à son interprétation.

La première que nous présentons est la structure thématique. La cohérence thématique concerne le thème d'un discours. La notion de thème est difficile à définir. Dans la langue courante, le thème est «ce sur quoi porte le texte dans son ensemble». En linguistique, les unités d'analyse sont plus petites : la phrase, la proposition ou l'unité syntaxique selon les approches. Le thème est donc considéré comme «ce sur quoi porte la phrase, la proposition ou l'unité syntaxique». La progression thématique d'un texte, envisagée du point de vue de la structure informationnelle dans la proposition ou la phrase, se manifeste par la présence de couples thème/rhème (Lambrecht, 1994). On s'intéresse ici, non pas au contenu des données, mais à la façon dont celles-ci sont présentées. Le thème correspond aux informations saillantes, données, présupposées ou connues. Il est le point de départ de l'énoncé, l'élément sur lequel on s'appuie pour apporter des informations nouvelles, i.e. le rhème. On parle alors de topique phrastique (pour le thème).

Cependant la structure thématique d'un texte dépasse largement le cadre de la phrase. Nous pouvons citer à titre d'illustration les circonstants thématiques qui jouent un rôle

important dans l'organisation globale du texte en assignant des thèmes à des segments de texte (Péry-Woodley, 2001) ou les *paragraphes thématiques* pour lesquels un thème, souvent introduit par la position grammaticale sujet, constitue un thème de niveau supérieur correspondant au sujet le plus impliqué dans la description (Givón, 1983). On parle alors de topique discursif.

Les notions de thème et de structure thématique permettent de rendre compte de certains choix syntaxiques et discursifs mais ne rendent pas compte à eux seuls de l'ensemble de l'organisation globale du texte.

La deuxième structure sous-jacente que nous évoquons est décrite au moyen de relations de discours, comme étant au centre de la cohérence. Il a été montré qu'un des aspects qui contribue à la cohérence d'un discours se manifeste par l'existence, entre les parties d'un texte, de liens, souvent nommés *relations de discours*, qui imposent une structure hiérarchique et forment la création de blocs.

Dans les approches qui étudient cette structure et ces relations de discours, on constate, dans les concepts utilisés, de nombreuses divergences et un manque de consensus. Les points de divergence les plus importants concernent la nature des segments, le nombre et la nature des relations de discours (*Cf.* Section 1.2).

Les méthodes d'analyse associées à ces modèles peuvent, par un phénomène de récursivité, procéder des plus petits segments définis (généralement les propositions) vers des unités plus grandes jusqu'au texte entier (analyse ascendante) ou décomposer le texte en unités plus grandes, jusqu'aux plus petits segments définis (analyse descendante).

Ces stratégies permettent de descendre du niveau global (texte dans son ensemble) au niveau local (unités de discours élémentaires) et inversement de grimper du niveau local au niveau global. Généralement, l'analyse procède soit de façon ascendante, soit de façon descendante.

La RST (Mann et Thompson, 1987) suppose que cette analyse peut procéder selon les deux approches. L'analyse de façon ascendante part du plus petit segment pour regrouper progressivement des segments de texte plus grands jusqu'à obtenir deux segments de texte reliés qui constituent la totalité du texte. L'analyse descendante va procéder à la segmentation de segments de texte plus larges (macro-segments), d'abord deux couvrant la totalité du texte, puis deux à deux jusqu'à l'obtention des unités de discours élémentaires (ou plus petits segments définis).

Le phénomène de récursivité descendante est mis en œuvre dans le modèle d'architecture textuelle (Luc, 2000). Le scripteur compose un texte T avec un titre t . T se décompose en parties $P_1, P_2, P_3 \dots P_n$ avec pour chacune «titre de P_1 », «titre de P_2 »... Chaque partie P se décompose en sous-parties $SP_1, SP_2, SP_3 \dots SP_n$ avec pour chacune «titre de SP_1 », «titre de SP_2 »... jusqu'à arriver aux paragraphes, plus petits segments délimités dans ce modèle. Ce modèle permet l'interprétation et la délimitation d'objets textuels qui sont ensuite mis en relation en faisant appel au modèle de la RST.

Le phénomène de récursivité ascendante est mis en œuvre dans le modèle de la SDRT. Les structures sont obtenues en construisant des représentations des segments complexes à partir de celles des segments simples. Dans ce processus de construction, un nouveau segment en cours de traitement doit être relié à un segment précédent par une relation de discours.

L'analyse dans le modèle de Grosz et Sidner (1986) procède par un imbriquement d'informations obtenues au niveau local et au niveau global. Grosz et Sidner (1986) définissent trois niveaux d'analyse, indépendants mais néanmoins reliés entre eux. La première composante concerne l'étude des énoncés (structure linguistique) et procède au niveau local. La deuxième composante est la structure intentionnelle. Elle procède également au niveau local en cherchant à déterminer le choix d'un contenu linguistique et d'un acte de langage en lien avec un but visé. Enfin la troisième composante rend compte du niveau global : la structure attentionnelle. Elle a pour objectif de distinguer les informations les plus saillantes des informations les moins saillantes. Ces trois niveaux d'analyse sont articulés les uns aux autres, ainsi les énoncés en cours de traitement remplissent-ils une fonction par rapport aux segments qui les contiennent (niveau local) et par rapport au discours entier (niveau global).

1.1.5 Bilan

La cohésion définit la qualité d'un texte tandis que la cohérence est à la fois un principe sous-tendant les processus interprétatifs et la qualité d'un discours bien construit.

Plus un texte est cohésif, plus la cohérence est obtenue sur la base d'éléments dépendants du texte. Lors de l'analyse d'un texte pour en construire une représentation, il s'agit, de notre point de vue, de tirer le maximum d'informations des indices linguistiques de la cohésion et de la progression thématique avant de faire appel aux aspects plus inférentiels de l'interprétation discursive.

1.2 Structures discursives et relations de discours

Les approches ascendantes qui cherchent à découvrir cette structure sous-jacente attribuée au discours et exploitable pour contribuer à son interprétation *via* l'existence de relations de discours posent les hypothèses suivantes :

- Tout discours cohérent a une structure.
- Il est possible de proposer une formalisation pour cette structure (bien que ce ne soit pas forcément l'objectif de toutes les approches).
- Cette structure permet de rendre compte de différents phénomènes discursifs (résolution anaphorique, structure temporelle...).
- Il existe un ensemble de relations de discours sur lesquelles repose l'interprétation du discours en un tout cohérent.

Si ces approches s'accordent sur le fait que le discours n'est pas linéaire tel que le texte le représente mais qu'il est structuré hiérarchiquement *via* des relations de discours reliant des segments de discours pour former des blocs, toutes ces approches ne s'accordent pas sur la nature des segments du discours, le nombre et la nature des relations de discours, la structure hiérarchique que ces relations imposent au discours et enfin sur la détection de ces relations de discours.

1.2.1 Nature des segments de discours

Les relations de discours s'établissent entre des segments de discours. Mais la nature exacte des segments de discours varie d'une approche à l'autre. Dans la RST et le modèle de Grosz et Sidner (1986), les segments de discours sont des unités intentionnelles (des segments auxquels on peut associer une intention, un but communicatif du scripteur/locuteur). Dans la SDRT et les modèles de Hobbs (1985) et de Polanyi (1988), les segments de discours sont des unités sémantiques (un segment dont le contenu sémantique correspond généralement à la description d'un état ou d'un événement).

Définir l'unité de discours élémentaire (ou UDE) et son identification joue un rôle crucial pour l'interprétation des discours. Pourtant, cette unité ne fait l'objet d'aucun consensus dans les modèles existants comme l'ont signalé Degand et Simon (2009). Il existe différentes conceptions du segment. Selon Grosz et Sidner, la phrase ne fait pas toujours l'objet d'un but communicatif et ce but peut-être véhiculé par un ensemble de phrases. Mais généralement, la phrase, ou la proposition, sont définies comme l'unité de discours élémentaire.

Si la nature exacte de l'UDE n'est pas clairement établie, les approches du discours s'accordent pour faire une distinction entre segment de discours simple (correspondant à une UDE) et segment de discours complexe (composé d'un groupement de plusieurs segments simples reliés par des relations de discours). Les segments complexes forment en quelque sorte l'emballage organisateur des segments simples. C'est par le biais de la construction de segments complexes que l'on peut atteindre une représentation de la globalité du texte.

1.2.2 Nombre des relations de discours

Concernant les relations de discours, on oppose les approches *réductionnistes* et les approches *multiplicatrices* suivant que la théorie propose un nombre réduit ou proliférant de relations de discours (Hovy et Maier, 1991).

La première approche, dite *réductionniste*, est typiquement illustrée par la théorie de Grosz et Sidner (1986) qui postulent que deux relations de discours sont suffisantes pour rendre compte de la structure de la totalité des discours. Selon eux, convenir d'un ensemble de relations plus détaillé est une tâche vouée à l'échec étant donné que plus on observe une relation de discours dans le détail, plus on va pouvoir la subdiviser en plusieurs autres relations de discours.

Deux principales critiques sont adressées à cette approche réductionniste de l'analyse des discours : le manque de précision linguistique et le manque de précision sémantique. Asher et Lascarides (2003) notent le manque de précision linguistique, qui leur ôte ainsi la possibilité de traiter les phénomènes spécifiquement linguistiques. Hovy et Maier (1991) montrent les limites de l'approche de Grosz et Sidner (1986) pour des applications de TAL, en donnant l'exemple de la génération de texte pour laquelle des relations plus sémantiques sont nécessaires.

Du côté des approches dites *multiplicatrices*, nous pouvons citer à titre d'exemple la RST et les travaux de Hovy et Maier (1991). On retrouve dans ces approches une distinction binaire entre relations coordonnantes et relations subordonnantes, qui rappelle les deux relations structurelles de Grosz et Sidner (1986). En revanche, dans la RST (Mann

et Thompson, 1988), relations coordonnantes (multi-noyaux) et relations subordonnantes (noyau-satellite)¹ sont subdivisées en 30 relations de discours.

Ce cadre permet, de plus, une analyse avec une liste ouverte de relations, c'est-à-dire que de nouvelles relations peuvent être ajoutées si cela paraît nécessaire lors de l'analyse d'un nouveau texte.

Hovy et Maier (1991) affinent l'éventail de relations de la RST en proposant d'organiser les relations de discours dans une hiérarchie de spécificités croissantes. Les points de départ de la hiérarchie sont les trois types de relations : relations de contenu, relations de présentation et relations textuelles. Puis chacun de ces ensembles est subdivisé en relations de discours plus spécifiques. La hiérarchie proposée par Hovy et Maier (1991) est ouverte, c'est-à-dire que les relations qui constituent les feuilles de la hiérarchie peuvent être à nouveau subdivisées en des relations plus spécifiques selon les besoins de l'analyse.

Le modèle proposé par la SDRT se base sur un ensemble fermé et réduit de relations de discours. Par exemple, en (1) et (2), une seule relation de discours d'*Explication* est inférée dans le cadre de la SDRT tandis que dans le cadre de la RST, le lecteur reconnaît en (1) une relation de *Cause non délibérée* et en (2) une relation de *Cause délibérée*.

(1) [Marie est tombée]_a [parce qu'elle a trébuché sur une marche.]_b

(2) [Marie a poussé Paul]_a [parce qu'il l'a insultée.]_b

La SDRT abandonne la contrainte posée par Polanyi (1988) et Mann et Thompson (1988) qu'une seule relation de discours peut relier deux segments. Dans l'exemple suivant, on peut appliquer la relation de *Contraste* déclenchée par le connecteur *mais* et la relation de *Narration* déclenchée par le connecteur *ensuite* :

(3) Jean a donné un livre à Marie, *mais* il le lui a *ensuite* repris (Busquets *et al.*, 2001, p. 81)

1.2.3 Nature des relations de discours

Pour définir l'ensemble des relations de discours, on distingue les approches basées sur la sémantique (Hobbs, 1985; Polanyi, 1988; Asher et Lascarides, 2003) de celles basées sur les intentions (Grosz et Sidner, 1986; Mann et Thompson, 1988). Dans les approches basées sur la sémantique, les relations de discours rendent compte des relations sémantiques qui peuvent exister entre des types d'états et d'événements. Dans les approches basées sur les intentions, les relations de discours rendent compte des buts communicatifs des scripteurs/locuteurs.

Les deux relations de discours structurelles de Grosz et Sidner (1986) sont établies au niveau de la structure intentionnelle : DOMINANCE est établie lorsqu'une action satisfait un but X et fournit une partie de la satisfaction d'une autre action Y, alors X contribue à Y et Y *dominates* X. SATISFACTION-PRECEDENCE est établie dans les cas où le but X a été satisfait avant le but Y, alors X *satisfaction-precedes* Y.

¹Nous reviendrons sur la structure hiérarchique qu'imposent les relations de discours dans la section ci-dessous 1.2.4.

Dans la RST, Mann et Thompson (1988) distinguent des *relations de contenu*, liées au contenu sémantique des propositions à relier et des *relations de présentation*, liées aux buts communicatifs. Mais toutes les relations sont définies selon les intentions du scripteur/locuteur : soit il a l'intention de signaler qu'il existe dans le monde une relation entre deux segments (relations de contenu), soit il cherche à influencer le point de vue de l'allocuteur ou lecteur sur le contenu du noyau (relations de présentation).

S'appuyant sur les travaux de Halliday (1985), Hovy et Maier (1991) rajoutent la catégorie des relations textuelles. Ces relations n'ont pas de sens sémantique ou intentionnel (*Conjonction*, *Pre-Séquence*...). Contrairement aux relations de contenu et aux relations de présentation qui font appel à des connaissances extérieures aux textes (faits sur le monde, informations liées au scripteur et lecteur impliqués dans le discours), les relations textuelles concernent des liens internes au texte. Les relations textuelles s'appliquent entre deux segments seulement si leur juxtaposition est imposée par la linéarité du texte. Elles sont typiquement utilisées pour énumérer une séquence d'arguments.

Dans le modèle de la SDRT, sont opposées les relations de contenu, liées au contenu sémantique des propositions à relier et les relations structurelles (*text structuring relations*), inférées sur la base d'un isomorphisme des structures ou la présence de marqueurs lexicaux spécifiques.

Revenons à nos deux exemples (1) et (2). Une approche utilisant des relations intentionnelles décrira la relation à l'œuvre *via* la reconnaissance d'une intention : en (1), le lecteur reconnaît dans le satellite la cause de l'action décrite dans le noyau et en (2), le lecteur reconnaît dans le satellite la cause de l'action délibérée décrite dans le noyau. Dans une approche sémantique, l'inférence de la relation d'*Explication* s'appuie sur des informations encodées dans les types des événements : l'événement décrit en (b) est un type d'événement étant une cause possible du type de l'événement décrit en (a).

1.2.4 Structure hiérarchique

La plupart des approches discursives s'accordent sur le fait que les relations de discours imposent une structure. À un point donné d'un discours, la contribution d'un nouveau segment en cours de traitement peut être de deux sortes : continuer ce qui est en cours dans le segment précédent ou le déterminer (expliquer, préciser...). Ces deux types de contribution sont qualifiés dans la plupart des théories du discours de *coordination* ou de *subordination*. Généralement, la coordination et la subordination sont représentées de façon arborescente par un trait horizontal pour les relations coordonnantes et un trait vertical pour les relations subordonnantes (Grosz et Sidner, 1986; Asher et Lascarides, 2003). On retrouve cette distinction dans (Hobbs, 1990) et dans la SDRT bien que toutes les relations, notamment celle d'*Élaboration*, ne soient pas toujours considérées comme imposant le même type de contrainte hiérarchique sur le discours : Hobbs (1990) considère la relation comme coordonnante et Asher et Lascarides (2003) comme subordonnante (Cf. Chapitres 2 et 3).

Dans la RST, Mann et Thompson (1987) distinguent les relations asymétriques des relations symétriques, distinction qui renvoie directement aux notions de coordination et subordination. Les relations asymétriques établissent une relation entre un noyau et un

(ou des) satellite(s), le satellite étant le segment de moindre importance. Une relation de dominance est ainsi appliquée entre le noyau et le satellite, tout comme le fait la subordination. Les relations symétriques établissent une relation entre deux noyaux, plaçant ces deux noyaux au même niveau d'importance, tout comme le fait la coordination.

1.2.5 Reconnaissance des relations de discours

Dans les approches basées sur les intentions (Grosz et Sidner, 1986; Mann et Thompson, 1987), la reconnaissance des relations de discours est établie par le biais de la reconnaissance des buts communicatifs du scripteur/locuteur. La RST (Mann et Thompson, 1987) propose un modèle descriptif des relations de discours, dans la perspective du *Recueil des Données* (Mann et Thompson, 2001) : la RST est essentiellement décrite comme une méthode pour recueillir des données. Cette démarche apparaît clairement à travers les types d'exemples qu'ils analysent, qui sont toujours des exemples attestés.

Par opposition à cette approche descriptive de la structure des textes, les approches sémantiques (Hobbs, 1990; Polanyi et Scha, 1984; Asher et Lascarides, 2003) cherchent à déterminer des règles d'inférence des relations de discours. Hobbs (1990) propose un raisonnement abductif, les inférences ainsi obtenues sont uniquement plausibles. Polanyi et Scha (1984) proposent des *grammaires de discours* qui cherchent à établir des analogies entre la composition des syntagmes dans une phrase et celle des énoncés dans un discours. Dans la SDRT (Asher et Lascarides, 2003), la reconnaissance des relations de discours est établie sur la base de la sémantique des énoncés individuels et des relations sémantiques qui existent entre les types d'état et d'événement pour atteindre compositionnellement le sens du discours. Les informations liées aux marqueurs lexicaux, structures syntaxiques, marqueurs aspectuo-temporels, sémantique lexicale et connaissances du monde et leurs interactions sont également pris en considération. La SDRT prend en compte ces différentes sources d'information et s'appuie sur une logique propositionnelle non monotone pour l'inférence des relations de discours et une logique pour représenter la sémantique des énoncés.

Par opposition à la RST qui défend fortement l'idée que les relations de discours sont accessibles en tenant compte uniquement des intentions et buts communicatifs des scripteurs/locuteurs, Knott (1996) défend, dans sa thèse, l'hypothèse que la reconnaissance des relations de discours peut se faire par le biais des marqueurs lexicaux, et uniquement par ce biais là. Nous reviendrons plus en détail sur la question de la signalisation des relations de discours dans le chapitre 4.

1.2.6 Positionnement

En considérant la relation de discours *Élaboration*, que nous présentons en détail dans les chapitres 2 et 3, nous nous interrogeons dans cette thèse, sur la signalisation des relations de discours (*Cf.* Chapitre 4). Nous n'essaierons pas de déterminer les intentions des scripteurs/locuteurs qui ont amené à la production de chaque segment de discours du fait que celles-ci ne sont généralement pas explicitement signalées dans les textes. Nous chercherons plutôt à déterminer les divers indices (marqueurs lexicaux, procédés syntaxiques, et connaissances lexicales et grammaticales sur les types d'éventualité et d'entité) et leurs

interactions sur lesquels se basent la reconnaissance des relations de discours dans une visée explicative en nous inscrivant dans le modèle théorique de la SDRT (qui accorde une place importante aux manifestations linguistiques des relations de discours). Cette approche sémantique implique de porter une attention particulière et détaillée aux réalisations linguistiques et à la complexité de leurs interactions. Cela entraîne nécessairement une limitation des données traitées. Étant donné que nous nous concentrons sur une seule relation de discours, cela va nous permettre de tester les descriptions proposées par le modèle sur un ensemble plus vaste de données attestées.

1.3 La SDRT

La SDRT est, comme la DRT dont elle constitue une extension, une théorie représentationnelle dynamique du discours. Elle ajoute à la DRT la prise en compte de la segmentation et de l'organisation structurelle du discours. Elle cherche à rendre compte des liens entre le contenu sémantique des propositions et la structure globale du discours. La SDRT, contrairement à la majorité des théories que nous avons abordées, est une théorie formelle qui décrit dans un cadre logique les mécanismes qui permettent aux lecteurs/allocutaires d'inférer une relation entre deux segments du discours.

Nous commencerons par présenter les principes généraux de la DRT, base théorique de la SDRT, qui vise à représenter le contenu sémantique des énoncés, puis nous présenterons la SDRT, qui vise à représenter la macro-structure du discours, en nous appuyant sur les synthèses de Busquets *et al.* (2001), Prévot (2004) et Bras (2008).

1.3.1 Représenter la sémantique du discours en DRT

La Théorie des Représentations Discursives (DRT - *Discourse Representation Theory*), introduite par Kamp (1981), vise la représentation du contenu sémantique du discours. La DRT est une approche logique de la sémantique des séquences d'énoncés. Elle propose d'en construire des représentations qui constituent un niveau intermédiaire entre la structure syntaxique et l'interprétation de ces représentations dans un modèle (représentation du monde suivant la tradition logique de la théorie des modèles).

1.3.1.1 Une théorie représentationnelle et dynamique

Les premiers travaux formels sur la signification proposent une représentation du sens des énoncés linguistiques en termes de conditions de vérité (Frege, 1892; Montague, 1974). Il s'agit d'une vision statique : les expressions linguistiques sont décrites selon la façon dont elles font référence à des objets (au sens large) du monde dont elles parlent. Ces approches sont limitées à la phrase et ne peuvent pas rendre compte des phénomènes de cohésion apparaissant au niveau discursif (anaphore interprastique, structure temporelle...)

La DRT s'appuie sur ces travaux et ajoute à la dimension référentielle statique une dimension interprétative dynamique. Les représentations, appelées DRS (*Discourse Representation Structure*), rendent compte du niveau interprétatif, à mi-chemin entre la structure syntaxique de la phrase et le modèle dans lequel l'énoncé est évalué en termes de condition de vérité relativement à une représentation du monde.

La représentation d'un discours est construite progressivement, au fur et à mesure de l'avancée de l'interprétation des séquences d'énoncés. Ce mode de construction rend compte de la nature incrémentale du processus d'interprétation (conception influencée par les approches procédurales de la psycholinguistique) et de la faculté des expressions à modifier le contexte. L'interprétation de chaque nouvelle expression linguistique met à jour un contexte discursif préalable (DRS construite jusque là) pour aboutir à un nouveau contexte discursif (nouvelle DRS mise à jour avec le contenu de l'énoncé en cours de traitement).

1.3.1.2 La représentation du discours

Dans l'analyse descendante proposée par Kamp et Reyle (1993), la représentation d'un discours est dérivée des arbres syntaxiques obtenus pour chacune des phrases. Le processus d'interprétation se fait en deux étapes : d'abord une analyse grammaticale classique et ensuite des structures déclenchantes qui permettent de passer d'une règle syntaxique à une règle de construction représentationnelle.

Nous allons illustrer le processus incrémental de l'interprétation en DRT avec l'analyse du mini-discours suivant et présenter en parallèle les propositions graphiques pour la représentation :

- (4) (a) Catherine mesure 157 cm. (b) Elle a les cheveux bruns et les yeux bleus. (c) Elle enseigne à l'université.²

La DRS est graphiquement représentée sous la forme d'une boîte, Cf. Figure 1.1. Une DRS \mathcal{K} est définie par la donnée d'un couple $\langle U_{\mathcal{K}}, C_{\mathcal{K}} \rangle$. La première partie de la boîte de la DRS contient l'ensemble des référents du discours, i.e. l'Univers $U_{\mathcal{K}}$. Et la deuxième partie contient l'ensemble des conditions, i.e. $C_{\mathcal{K}}$, autrement dit l'ensemble des spécifications sur les référents du discours et les relations qui les relient.

Une DRS se construit de façon incrémentale de la façon suivante : le point de départ est la DRS \mathcal{K}_0 , qui représente le contexte vide. Le traitement de la première phrase (a) produit la DRS \mathcal{K}_a . Cette DRS sert alors de contexte pour le traitement de la seconde phrase (b) pour aboutir à \mathcal{K}_b . Enfin, la DRS \mathcal{K}_b va servir de contexte pour le traitement de la dernière phrase (c) pour aboutir à la DRS finale \mathcal{K}_c (Figure 1.1).

Chaque expression linguistique de chaque phrase apporte sa propre contribution lors de la construction de la DRS.

Pour la phrase (a) de l'exemple (4), le nom propre *Catherine* permet d'introduire un référent de discours, nommé v et une condition $Catherine(v)$ qui rattache le référent v au nom propre *Catherine*. Le référent s_1 est un référent d'état. Les éventualités, événements et états, sont représentées par un référent de discours, suivant les propositions de Davidson (1967). Le contenu propositionnel d'un état comme «mesurer 157 cm» est représenté par un prédicat à deux arguments $s_1 - \text{mesurer } 157 \text{ cm}(v)$ qui est la notation employée dans (Asher, 1993) ou $\text{mesurer } 157 \text{ cm}(s_1, v)$ qui est une notation équivalente. Enfin dans cette DRS, une constante temporelle n est introduite, qui représente le moment de la parole *now*. Ces deux référents temporels, référent d'état et constante temporelle, sont mis en relation dans une condition temporelle de chevauchement, notée \circ .

²Cet exemple est traduit de (Asher, 1993, p. 276). Il sera repris Section 3.1.1 lors de la présentation de la relation d'*Élaboration*.

La DRS \mathcal{K}_a va servir de contexte discursif au traitement de la phrase (b). La seconde phrase (b) de (4) contient un pronom anaphorique «elle», représenté par z . Pour résoudre cette anaphore ($z = ?$, appelé équation anaphorique), il est nécessaire de mettre le pronom anaphorique en relation avec un référent de discours déjà présent dans le contexte, i.e. dans l'Univers de la DRS \mathcal{K}_a . Cette relation est exprimée par la condition $z = v$ après résolution. Cette équation permet de mettre en relation le nouveau référent de discours introduit par le pronom avec un référent de discours déjà introduit dans le contexte de la DRS. Les référents s_2 et s_3 sont des référents temporels d'état, ici représentés par des prédicats à trois arguments. L'état «Elle a les cheveux bruns» est représenté par $s_2 - avoir(z, x)$, z est le référent de discours introduit par «elle» et x est le référent de discours introduit par la forme plurielle «les cheveux bruns» mis en relation par la condition $cheveux - bruns(x)$. Nous pouvons suivre le même raisonnement pour le deuxième référent temporel d'état s_3 : l'état «elle a les yeux bleus» est représenté par $s_3 - avoir(z, y)$ où z est le référent introduit par «elle» et y le référent introduit par la forme plurielle «les yeux bleus» mis en relation par la condition $yeux - bleus(y)$. Enfin, chacun des deux référents temporels d'état est mis en relation dans une condition temporelle de chevauchement avec le moment de la parole n .

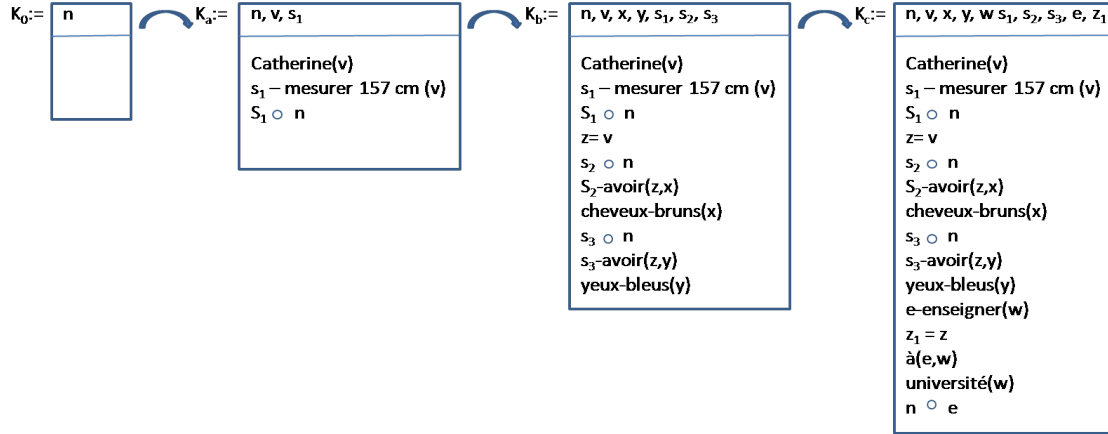
La DRS \mathcal{K}_b va alors servir de contexte discursif au traitement de la phrase (c). La troisième phrase (c) contient également un pronom anaphorique «elle», représenté par z_1 . Le pronom anaphorique est mis en relation avec un référent de discours déjà présent dans le contexte discursif, ici avec z . Cette relation est exprimée par la condition $z_1 = z$. Le référent e est un référent temporel d'événement. L'événement³ «Elle enseigne à l'université» est représenté par un prédicat à deux arguments $e - enseigner(z_1)$, z_1 est le référent de discours introduit par «elle». La préposition «à» est considérée comme mettant en place un prédicat à trois arguments : $\hat{a}(e, w)$, e est l'événement et w le référent de discours introduit par la forme définie «l'université» mis en relation par la condition $université(w)$. Enfin, le référent temporel d'événement e est mis en relation dans une condition temporelle de chevauchement avec le moment de la parole n .

Les éventualités présentes en (4) sont soit des états, soit des événements, selon la manière dont la langue encode celles-ci. La SDRT s'appuie sur la distinction entre état et événement proposée par Kamp et Reyle (1993) pour lesquels état et événement se caractérisent par leur fonctionnement dans les textes narratifs. Les événements assurent la progression narrative tandis que les états fournissent des informations d'arrière-plan.

1.3.2 Représenter la structure du discours en SDRT

La SDRT (Segmented Discourse Representation Theory) a cherché à réunir les travaux de la DRT qui analyse le contenu propositionnel du discours et les travaux de l'analyse du discours. C'est donc une théorie représentationnelle dynamique qui prend en compte la segmentation et l'organisation structurale du discours *via* les relations de discours qui s'établissent entre les segments. Cette extension a été motivée par des problèmes empiriques non résolus par la DRT tels que l'anaphore interphrastique et certains problèmes liés à la structure temporelle des discours.

³ «Elle enseigne à l'université» est traité comme un événement dans Asher (1993).

FIG. 1.1 – DRS \mathcal{K}_c de la phrase (c)

1.3.2.1 Unités de discours élémentaires et relations de discours

Unités de discours élémentaires et segments complexes En SDRT, le discours est envisagé comme un ensemble de segments simples ou complexes reliés entre eux par des relations de discours. Les unités de discours élémentaires sont généralement la proposition syntaxique ou la phrase. Elles sont représentées par des DRS jouant le rôle de constituants élémentaires. Les relations de discours vont regrouper plusieurs unités de discours élémentaires de sorte à former des constituants complexes, qui sont alors représentés par des SDRS. Les SDRS sont ainsi des structures récursives.

Relations de discours Les relations en SDRT sont fondées sur des critères sémantiques. Elles ont des effets sur les conditions de vérité du contenu propositionnel des segments reliés. Une nouvelle relation peut être définie uniquement si les effets sémantiques de celle-ci sont distincts des effets sémantiques des relations de discours déjà existantes.

Comme toutes les théories du discours basées sur des relations de discours, la SDRT développe également les contraintes hiérarchiques posées par les relations de discours, i.e. leur capacité à organiser le discours en groupant des segments minimaux et en déterminant des fonctions pour ces groupes de segment.

Deux relations de discours peuvent être vérifiées entre deux segments de discours. Cela est possible à condition que les effets sémantiques des deux relations en question soient compatibles.

Pour qu'un texte soit considéré comme cohérent, il faut que toutes ses parties soient attachées à une autre par au moins une relation de discours.

Asher et Lascarides (2003) signalent que la cohérence n'est pas une propriété binaire, mais qu'au contraire la qualité de la cohérence peut varier. Ils décrivent un modèle pour rendre compte d'une cohérence «scalair» basée sur trois règles :

1. Plus il existe de connections rhétoriques entre deux constituants d'un discours, plus

l'interprétation est cohérente.

2. Plus les expressions anaphoriques sont résolues, meilleure est la qualité de la cohérence de l'interprétation.

3. Certaines relations rhétoriques sont elles-mêmes scalaires. Par exemple, la qualité d'une *Narration* dépend de la spécificité du topique commun qui résume les segments de l'histoire ; et la qualité d'un *Contraste* est dépendante des éléments contenus dans les segments qui sont en relation. Une interprétation qui maximise la qualité des relations rhétoriques est plus cohérente qu'une qui ne le fait pas.

Dans la version de la SDRT de (Asher et Lascarides, 2003), une liste de relations de discours est proposée (par ordre alphabétique) : *Alternation*, *Arrière-plan*, *Commentaire*, *Conséquence*, *Continuation*, *Contraste*, *Élaboration*, *Explication*, *Narration*, *Parallèle*, *Précondition*, *Résultat*, *Topique*⁴.

La structure hiérarchique Les relations de discours posent des contraintes sur la représentation du discours. L'ensemble des relations de la SDRT se divise en relations coordonnantes et subordonnantes.

Les relations subordonnantes introduisent une hiérarchie entre deux segments. Un des deux segments est dominé par l'autre. La relation d'*Élaboration* a été la première relation considérée comme subordonnante.

Les relations coordonnantes placent les deux segments au même niveau. Les deux segments ainsi reliés forment alors un nouveau segment complexe et se trouvent automatiquement dominés par un constituant, appelé *topique discursif*. Si ce topique n'est pas donné dans le discours, il faut alors construire le contenu sémantique de ce topique à partir du contenu des énoncés qu'il domine et de la relation qui les lie (sauf pour *Résultat*). On parle dans ce cas de *topique implicite*.

La hiérarchisation du discours obtenue par la distinction des relations coordonnantes et subordonnantes permet de définir, parmi les constituants reliés par des relations de discours, ceux qui sont disponibles comme point d'attachement pour la poursuite du discours. Les constituants ou sites disponibles pour la poursuite du discours sont dits *ouverts* (par opposition à *fermés*). Ils se situent sur la *frontière droite*, Cf. Figure 1.2.

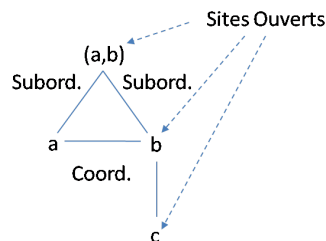


FIG. 1.2 – Illustration de la frontière droite

En plus de définir les sites d'attachement disponibles pour la poursuite du discours,

⁴Nous ne listons pas les relations de discours spécifiques au dialogue.

la frontière droite permet également de déterminer l'accessibilité des référents pour la résolution anaphorique. Seuls les référents introduits dans des constituants sur la frontière droite sont accessibles pour la résolution anaphorique.

Les constituants situés à droite du schéma sont les constituants ouverts. L'attachement d'un constituant par une relation coordonnante à un autre constituant a pour effet de rendre ce constituant inaccessible pour l'attachement d'un nouveau constituant et pour la résolution anaphorique.

La reconnaissance des relations de discours La reconnaissance des relations de discours, comme nous l'avons déjà mentionné, est établie sur la base de la sémantique des énoncés individuels, des marqueurs lexicaux, des structures syntaxiques, des marqueurs aspectuo-temporels, de la structure lexicale et des connaissances du monde. La SDRT est en cela une théorie de l'interface sémantique/pragmatique.

Asher et Lascarides (2003) accordent au contexte une place importante dans l'interprétation des discours. Le contexte étroit est l'ensemble des informations nécessaires (identité du locuteur, environnement spatio-temporels) pour résoudre certaines expressions référentielles (i.e. les indices linguistiques de la cohésion) et le contexte large est tout ce que le lecteur ou l'allocutaire va utiliser pour déterminer les intentions communicatives du scripteur ou locuteur.

La SDRT intègre toutes ces informations qui interviennent dans le processus d'interprétation du discours dans un but opératoire dans la construction des représentations du discours au moyen de règles de déclenchement des relations de discours.

1.3.2.2 Construction des SDRS

Les constituants élémentaires ou DRS sont reliés entre eux par des relations de discours pour former des constituants complexes représentés au moyen de SDRS. Dans les premières versions de la SDRT (Asher, 1993), les relations de discours relient directement les constituants (représentation du contenu propositionnel) entre eux. Dans les versions plus récentes (Asher et Lascarides, 2003), les relations de discours relient les étiquettes des constituants⁵. Les étiquettes sont généralement représentées dans les schémas par des lettres grecques ($\alpha, \beta, \gamma, \sigma$). On note \mathcal{K}_α une DRS étiquetée par α et e_α l'éventualité décrite par \mathcal{K}_α . Le constituant ou énoncé est noté $\alpha : \mathcal{K}_\alpha$. Une relation de discours entre deux étiquettes est notée $R(\alpha, \beta)$.

Une SDRS peut être graphiquement représentée sous la forme d'une boîte (tout comme les DRS). Une SDRS est définie par la donnée d'un couple $\langle U, C \rangle$ tel que U est un ensemble d'étiquettes et C un ensemble de conditions. Les étiquettes renvoient à des constituants simples ou à des constituants complexes. Les conditions sont soit la représentation d'une instance de proposition $\alpha : \mathcal{K}_\alpha$, soit la représentation d'une relation $R(\alpha, \beta)$.

Nous allons illustrer la construction d'une SDRS à partir des exemples suivants empruntés à Bras (2008, p. 33) :

- (5) (a) Marie a appris une poésie. (b) Puis elle a fait un exercice.

⁵Les étiquettes correspondent à la trace d'un acte de langage.

- (6) (0) Marie a fait ses devoirs. (a) Elle a appris une poésie. (b) Puis elle a fait un exercice.

Si l'on admet que les deux segments de l'exemple (5) sont reliés par la relation de *Narration*, nous obtenons la SDRS en Figure 1.3 où U est l'ensemble des étiquettes π_a et π_b dont les DRS sont données, et C est l'ensemble des trois conditions $\pi_a : \mathcal{K}_{\pi_a}$, $\pi_b : \mathcal{K}_{\pi_b}$, $Narration(\pi_a, \pi_b)$.

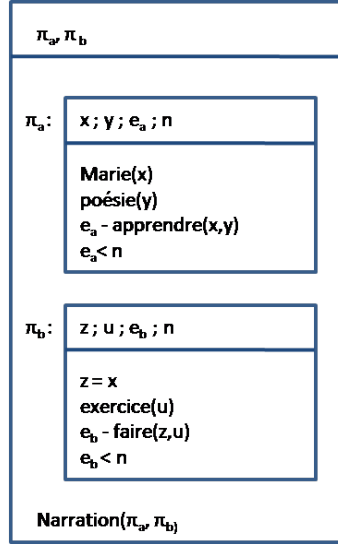


FIG. 1.3 – SDRS de (5)

Dans l'exemple (6), π_a et π_b forment ensemble le constituant complexe étiqueté π' , qui élabore le constituant π_0 en Figure 1.4.

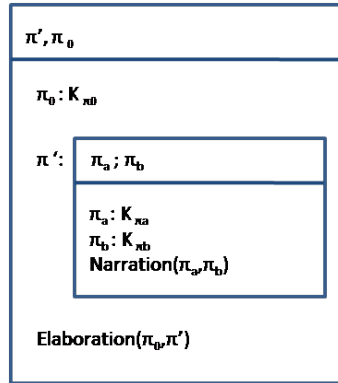


FIG. 1.4 – SDRS de (6)

Les SDRS peuvent être représentées sous forme de graphique (Cf. Figure 1.5), qui tout en occultant le contenu sémantique de chaque DRS, met en évidence la structure hiérarchique

avec des traits verticaux pour les relations subordonnantes et horizontaux pour les relations coordonnantes.

Nous optons pour la représentation graphique (celle de droite) qui utilise des boîtes pour les constituants complexes. Ces types de graphique seront notamment employés dans le projet ANNODIS car ceux-ci représentent, de façon simplifiée, les constituants complexes sous la forme d'un cadre.

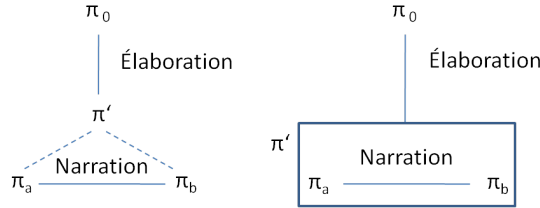


FIG. 1.5 – Graphiques SDRS de (6)

Les SDRS, tout comme les DRS, sont construites de façon incrémentale par le traitement une à une des UDE suivant trois étapes principales :

La première étape consiste à associer à chaque UDE une DRS. Suivant le modèle de la DRT, la DRS est construite à partir de l'analyse syntaxique de l'unité en cours de traitement.

La deuxième étape consiste à déterminer les sites ouverts pour l'attachement de l'étiquette du constituant en cours de traitement. Contrairement à la DRS, la SDRS est une structure hiérarchisée. L'étiquette associée à un constituant peut donc être attachée en divers points de la structure définis par la frontière droite. Les sites ouverts (ou la frontière droite) correspondent au dernier constituant introduit et à tous les constituants qui le dominent hiérarchiquement.

Pour tous les sites accessibles, il faut déterminer la ou les relations de discours qui existe(nt) pouvant réaliser l'attachement de la nouvelle étiquette. Les relations de discours sont inférées, comme nous l'avons déjà mentionné, en s'appuyant sur des informations de diverse nature. Tous ces indices relevés qui jouent un rôle dans l'interprétation d'une relation de discours sont partiels et généralement ambigus, la SDRT a donc fait appel à des mécanismes de raisonnement non monotone, permettant d'exprimer des règles révisables. Les règles peuvent être exprimées avec l'opérateur \rightarrow (pour les règles monotones, inférées en dur et non révisable) et \succ (pour les règles non monotones révisables). Ces règles sont appelées *règles de déclenchement des relations de discours*. D'autres règles sont ensuite associées aux relations de discours, ce sont les règles formalisant les *effets sémantiques des relations de discours*. Ces règles expriment des conséquences discursives sur les constituants reliés ou sémantiques sur le contenu des constituants reliés.

Dans la section 1.3.2.3, nous donnerons les descriptions (règles de déclenchement et effets sémantiques) des relations de discours que nous utiliserons dans nos futures analyses.

Enfin, la dernière étape de la construction d'une SDRS consiste en sa mise à jour. La mise à jour se fait une fois que sont déterminés le site d'attachement et la ou les relation(s) de discours, en deux temps : d'abord la mise à jour de la structure avec les constituants complexes et les topiques implicites le cas échéant ; ensuite une mise à jour du contenu des constituants avec la résolution des sous-spécifications.

Une illustration de la construction d'une SDRS est donnée dans le chapitre 3 (section 3.1.3.3) consacré à la description de la relation d'*Élaboration* dans la SDRT.

L'interprétation d'un discours en SDRT est conditionnée par le succès de l'attachement d'une nouvelle étiquette à un site ouvert et l'inférence d'une relation de discours pour réaliser cet attachement. Si cette construction échoue, le discours est déclaré incohérent.

Pour qu'une SDRS soit cohérente, toute étiquette d'un constituant doit être reliée à une autre étiquette de la SDRS par au moins une relation de discours. L'interprétation d'un discours en SDRT peut aboutir à plusieurs SDRS cohérentes. La théorie fait alors appel à un principe, appelé *Maximise Discourse Coherence* (ci-après MDC) qui ordonne les SDRS selon le nombre de relations de discours qui relient les étiquettes. La SDRS comportant le maximum de relations de discours, le minimum de noeuds entre étiquettes et les meilleures résolutions sera considérée comme la représentation la plus cohérente du discours.

1.3.2.3 Inférence des relations de discours

Dans cette section, nous allons décrire les relations de discours que nous emploierons dans nos analyses au sein de cette thèse, sauf celles d'*Élaboration* et de *Continuation* dont les descriptions seront données au chapitre 3, en nous appuyant sur l'ouvrage de Asher et Lascarides (2003). Toutes ces relations, excepté *Alternation*, sont des relations *véridiques*, c'est-à-dire que le nouveau contexte qu'elles établissent satisfait le contenu propositionnel des constituants reliés.

Les relations de discours sont décrites en deux temps. Les règles de déclenchement, d'une part, indiquent sous forme d'axiome de la Glue Logic la façon dont on infère les relations de discours. Ces règles utilisent un méta-langage qui manipule des SDRS. À gauche de la règle, on trouve un prédicat à trois places $R(\alpha, \beta, \lambda)$ avec R pour relation, α et β pour les étiquettes et λ le constituant minimal qui contiendra α et β . $R(\alpha, \beta, \lambda)$ indique qu' α et β sont reliés par une relation R dans la SDRS λ . $?(\alpha, \beta, \lambda)$ signifie qu'il existe une relation, non encore déterminée, entre α et β dans la SDRS λ . À droite de la règle, on trouve des formules s'appuyant sur un marqueur du contenu propositionnel des constituants, noté $[marqueur]\alpha$, une structure syntaxique et/ou un prédicat spécifique.

Les effets sémantiques des relations de discours sont donnés dans un second temps. Le rôle des effets sémantiques est double. Ces effets définissent les conséquences des relations et des contraintes sémantiques qui viennent enrichir le contenu des constituants reliés. Au niveau du *Maximise Discourse Coherence*, il est assuré que deux relations de discours avec des effets sémantiques incompatibles ne soit pas inférées en même temps.

Dans (Asher et Lascarides, 2003), ces règles sont écrites dans le langage du contenu informationnel des SDRS.

Nous ne visons pas l'exhaustivité pour les descriptions de ces relations. Nous nous limiterons aux axiomes majeurs que nous emploierons dans nos analyses.

Les premières définitions que nous fournirons sont celles des relations dites «de contenu» (*Content-Level Relations*). Les relations de ce type sont entièrement définies sur la base de la sémantique des éventualités et individus introduits dans les SDRS.

Narration Deux segments reliés par la relation de *Narration* décrivent deux événements d'une même histoire qui se sont déroulés successivement. La relation de *Narration* est une relation coordonnante. Nous illustrons cette relation avec les exemples de (Bras, 2008, p. 40).

- (7) Pierre frappa à la porte. Il entra.
- (8) Pierre tomba. Puis Paul le poussa.

La relation de *Narration* peut être inférée entre deux segments en s'appuyant sur la présence du marqueur spécifique «Puis» comme dans (8). Cette inférence est rendue possible par la règle d'inférence monotone suivante (Bras *et al.*, 2001, p. 123) :

Axiome 1.1 Inférer Narration Puis $(?(\alpha, \beta, \lambda) \wedge [puis](\beta)) \rightarrow Narration(\alpha, \beta)$

La relation de *Narration* peut également être inférée sur la base d'un prédicat spécifique appelé *Occasion*, appliqué aux types des éventualités décrites dans les constituants à relier. Ce prédicat est vérifié si les types des éventualités décrites dans les deux segments appartiennent à une même histoire, sur la base d'informations de sémantique lexicale et de connaissances du monde. Pour l'exemple (7), les types des éventualités «frapper à la porte» et «entrer» peuvent appartenir à la même histoire, l'un arrivant généralement après l'autre, nous amenant à l'inférence de *Narration* (Bras *et al.*, 2001, p. 120).

Axiome 1.2 Inférer Narration Occasion
 $(?(\alpha, \beta, \lambda) \wedge Occasion(\alpha, \beta)) > Narration(\alpha, \beta)$

Dans les premières versions de la SDRT *Narration* est considérée comme la relation de discours par défaut, se déclenchant alors dans tous les cas mais pouvant être révisée par une règle plus spécifique. Avec l'introduction de la règle basée sur le prédicat *Occasion*, une nouvelle règle est introduite spécifiant que *Narration* est inférée si et seulement s'il n'existe pas d'indices d'autres relations de discours (Bras *et al.*, 2001, p. 120).

Axiome 1.3 Inférer Narration Défaut
 $?(\alpha, \beta, \lambda) \wedge \neg clues_R_1(\alpha, \beta, \lambda) \wedge \neg clues_R_2(\alpha, \beta, \lambda) \wedge \dots \wedge \neg clues_R_n(\alpha, \beta, \lambda) > Narration(\alpha, \beta)$

La relation de *Narration* impose des contraintes temporelles sur le contenu propositionnel des constituants reliés. Les éventualités décrites par les constituants reliés doivent être ordonnées temporellement (Lascarides et Asher, 1993) :

Axiome 1.4 Narration Conséquence Temporelle $\phi_{Narration(\alpha, \beta)} \Rightarrow e_\alpha \prec e_\beta$

Cette relation impose également des contraintes structurelles : deux constituants reliés par *Narration* doivent avoir un topique commun, i.e. une sorte de résumé du contenu propositionnel des deux constituants. Ce topique peut être explicite et déjà introduit dans la structure du discours comme dans le cas de (6).

Mais si ce topique n'est pas introduit, alors une première règle entraîne la construction d'un topique implicite au moyen de l'opérateur (\sqcap) qui calcule le contenu commun des constituants reliés. L'opérateur modal de nécessité (\Box) spécifie que ce topique ne peut pas être vide de sens (Asher et Lascarides, 2003, p. 164).

Axiome 1.5 *Narration Contrainte Topique*

$$\phi_{Narration(\alpha, \beta)} \Rightarrow \neg \Box(\mathcal{K}_\alpha \sqcap \mathcal{K}_\beta)$$

Une règle intervient ensuite au moment de la mise à jour de la SDRS, par l'ajout d'un constituant implicite et d'un constituant complexe formé des deux étiquettes reliées par *Narration*. Le constituant complexe est alors dominé par le constituant implicite au moyen de la relation de Topique, notée \Downarrow .

Axiome 1.6 *Narration Contrainte Topique MAJ*

$$Narration(\alpha, \beta, \lambda) \rightarrow \exists \delta((\delta = \alpha \sqcap \beta) \wedge \exists \gamma(\Downarrow(\gamma, \lambda, \delta)))$$

Résultat Deux segments sont reliés par la relation de *Résultat* si le second segment résulte du premier, comme dans les exemples typiques de la relation :

(9) Marie a poussé Pierre. Il est tombé.

(10) Marie a éteint les lumières. Il faisait nuit noire autour d'elle.

La relation de *Résultat* peut être inférée de façon non monotone sur la base d'un prédicat spécifique appelé *Cause_D*, appliqué aux types des éventualités décrites dans les constituants à relier. Ce prédicat est vérifié si le type de l'éventualité décrite dans α est une cause possible du type de celle décrite dans β (l'axiome restreint la portée de *Cause_D* à σ pour lequel *Top* signale qu'il est le sommet de la structure discursive dans laquelle apparaît α) dans le discours D , sur la base d'informations générales de sémantique lexicale et de connaissances du monde (Bras *et al.*, 2001, p. 121).

Axiome 1.7 *Inférer Résultat*

$$(?(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge Cause_D(\sigma, \alpha, \beta)) > Résultat(\alpha, \beta, \lambda)$$

Par exemple pour (9), des informations sur la sémantique des types de verbe «pousser» et «tomber» permettent de vérifier le prédicat *Cause_D* entre les deux éventualités.

La relation de *Résultat* peut également être inférée grâce à la présence du marqueur *donc* (Bras *et al.*, 2001, p. 121) :

Axiome 1.8 *Inférer Résultat Donc* $(?(\alpha, \beta, \lambda) \wedge [donc](\beta)) \rightarrow Résultat(\alpha, \beta)$

Dans la version de (Asher et Lascarides, 2003), la relation de *Résultat* est considérée comme une relation coordonnante. Mais des travaux récents ont mis en évidence la nature tantôt coordonnante, tantôt subordonnante de la relation (Asher et Vieu, 2005).

La relation de *Résultat* impose un lien causal entre les éventualités. Ce lien causal impose ensuite que les événements décrits par les constituants reliés par la relation se succèdent dans le temps (Bras *et al.*, 2001, p. 121).

Axiome 1.9 *Résultat Conséquence* $\phi_{Résultat(\alpha,\beta)} \Rightarrow Cause(e_\alpha, e_\beta)$
 $(cause(e_\alpha, e_\beta) \wedge event(e_\beta)) \Rightarrow e_\alpha \prec e_\beta$

Explication La relation d'*Explication* est la relation inverse de la relation de *Résultat*. Deux segments sont reliés par la relation d'*Explication* si le second segment explique le premier, comme dans les exemples typiques de la relation :

(11) Pierre est tombé. Marie l'a poussé.

Explication est une relation subordonnante pour laquelle l'étiquette du segment expliqué domine celle du segment expliquant. La relation d'*Explication* peut être inférée sur la base du prédicat $Cause_D$, appliqué aux types des éventualités décrites dans les constituants à relier. Ce prédicat est vérifié si le type de la seconde éventualité est une explication possible du type de la première, sur la base d'informations générales de sémantique lexicale et de connaissances du monde. L'axiome de déclenchement non monotone de la relation d'*Explication* est symétrique à l'axiome de déclenchement de *Résultat* (Bras *et al.*, 2001, p. 122).

Axiome 1.10 *Inférer Explication*
 $(?(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge Cause_D(\sigma, \beta, \alpha)) \succ Explication(\alpha, \beta, \lambda)$

La relation d'*Explication* peut également être inférée grâce à la présence du marqueur *car* (Bras, 2008, p. 43) :

Axiome 1.11 *Inférer Explication Car* $(?(\alpha, \beta, \lambda) \wedge [car](\beta)) \rightarrow Explication(\alpha, \beta)$

Les effets sémantiques de la relation sont également symétriques à ceux de *Résultat*. La relation d'*Explication* impose un lien causal entre les éventualités. Ce lien causal impose ensuite que les événements décrits par les constituants reliés par la relation se succèdent dans le temps (Asher et Lascarides, 2003).

Axiome 1.12 *Explication Conséquence* $\phi_{Explication(\alpha,\beta)} \Rightarrow Cause(e_\beta, e_\alpha)$
 $(cause(e_\alpha, e_\beta) \wedge event(e_\beta)) \Rightarrow e_\alpha \prec e_\beta$

Arrière-plan Deux segments sont reliés par la relation d'*Arrière-plan* si le second segment donne des informations relatives à l'arrière-plan (donné sous forme d'état) de l'événement décrit dans le premier segment. On dit que cet événement est situé à l'avant-plan.

(12) Paul entra. Marie faisait la vaisselle.

(13) Marie faisait la vaisselle. Paul entra.

Pour inférer la relation d'*Arrière-plan* entre les étiquettes des deux constituants, il faut qu'un des constituants décrive un événement tandis que l'autre décrit un état, peu importe l'ordre comme le montrent les deux exemples ci-dessus. La relation d'*Arrière-plan* est alors vérifiée, respectivement au moyen de ces deux règles d'inférence (Asher et Lascarides, 2003).

Axiome 1.13 Inférer Arrière-plan1

$$(?(\alpha, \beta, \lambda) \wedge event(\alpha) \wedge state(\beta) > Arrière-plan(\alpha, \beta, \lambda)$$
Axiome 1.14 Inférer Arrière-plan2

$$(?(\alpha, \beta, \lambda) \wedge state(\alpha) \wedge event(\beta) > Arrière-plan(\beta, \alpha, \lambda)$$

La relation d'*Arrière-plan* impose des contraintes temporelles sur le contenu propositionnel des constituants reliés. Les éventualités décrites par les constituants reliés doivent se chevaucher temporellement (Asher et Lascarides, 2003) :

Axiome 1.15 Arrière-plan Conséquence Temporelle $\phi_{Arrière-Plan(\alpha, \beta)} \Rightarrow \alpha \circ \beta$

Dans la version de (Asher et Lascarides, 2003), la relation d'*Arrière-Plan* est une relation subordonnante. Deux relations sont distinguées selon que l'arrière-plan est donné en premier ou en second, Cf. les exemples (12) et (13).

Précondition Cette relation de discours a été introduite dans (Bras et Asher, 1994) pour rendre compte des retours en arrière dans les textes narratifs. La relation de *Précondition* est inférée lorsque qu'un segment introduit une éventualité qui correspond à un retour en arrière dans la narration. L'inférence de cette relation s'appuie sur des indices grammaticaux (emploi du plus-que-parfait) ou sur des informations alliant connaissances du monde et sémantique lexicale.

Alternation La relation d'*Alternation* est équivalente à la disjonction de deux constituants. C'est une relation non véridique, c'est-à-dire que les constituants reliés par *Alternation* transforment le contexte en un contexte qui satisfait uniquement le contenu propositionnel d'un des deux constituants reliés.

Les deux prochaines relations, *Contraste* et *Parallèle* que nous allons décrire ci-dessous, sont des relations structurantes au niveau textuel (*Text Structuring Relations*). Ces relations ne sont pas définies sur la base du contenu sémantique des constituants comme c'est le cas pour les relations de contenu mais par un isomorphisme des structures des constituants et/ou la présence d'un marqueur spécifique.

Contraste La description de la relation de *Contraste* repose sur un isomorphisme structurel des constituants et sur la notion de thème contrastif. L'emploi de verbes avec des polarités opposées ou de verbes avec des sémantismes opposés, configurations respectivement illustrées en (14) et (15), permet de marquer des thèmes contrastifs :

(14) John aime un peu le sport. Mais Pierre aime beaucoup.

(15) John aime le sport. Mais il hait le football.

La relation de *Contraste* peut être inférée entre deux constituants de façon non monotone en présence des indices évoqués ci-dessus : l'isomorphisme structurel et la notion de thème contrastif (Bras *et al.*, 2001, p. 123).

Axiome 1.16 Inférer Contraste

$$(?(\alpha, \beta, \lambda) \wedge \text{Structurally_similar}(\alpha, \beta) \wedge \text{Contrasting_themes}(\alpha, \beta)) > \text{Contraste}(\alpha, \beta, \lambda)$$

La relation de *Contraste* peut également être inférée de façon monotone en présence d'un marqueur explicite comme *mais* (Bras *et al.*, 2001, p. 123) :

Axiome 1.17 Inférer Contraste Mais $(?(\alpha, \beta, \lambda) \wedge [\text{mais}](\beta) \rightarrow \text{Contraste}(\alpha, \beta, \lambda)$

La relation de *Contraste* n'impose pas de contraintes temporelles sur le contenu des constituants reliés. *Contraste*(α, β) impose des contraintes sur les implicatures de α et β . Les faits impliqués par le contenu propositionnel de α et le contenu propositionnel de β sont incompatibles.

Parallèle La description de la relation de *Parallèle* est similaire à celle de contraste. Elle repose également sur un isomorphisme des structures des constituants. En revanche, elle fait intervenir la notion de thème commun partagé par les constituants (Asher *et al.*, 2001).

1.4 Structures discursives et applications de TAL

Ce chapitre a été largement consacré jusqu'ici à la présentation de diverses approches ayant pour objectif de décrire ou de modéliser la structure du discours *via* des relations de discours. Nous souhaitons présenter maintenant des approches appliquées qui visent à manipuler et à détecter automatiquement des éléments de cette structure discursive étudiée par les approches fonctionnelles et formelles. Ces approches appliquées visent l'enrichissement et l'approfondissement d'applications de TAL telles que le résumé automatique, la génération de texte avec des informations tirées de la structure discursive.

La description et la détection de la structure du discours ont fait émerger une quantité de travaux dont nous ne pouvons pas faire un état de l'art exhaustif. Dans cette section, nous souhaitons présenter un état de la recherche concis concernant les structures discursives et les applications en TAL.

1.4.1 Distance entre approches théoriques et applications de TAL

La synthèse de Moore et Wiemer-Hastings (2003) fait état d'un écart entre les approches théoriques de la structure du discours qui décrivent et modélisent cette dernière et les applications qui tentent de détecter des éléments de cette structure automatiquement. Moore & Wiemer-Hastings constatent que les applications telles que le résumé automatique ou les systèmes de question-réponse, qui prennent en compte des informations liées à la structure discursive, ont rompu avec les théories de la structure du discours, laissant place aux performances statistiques et mémorielles des systèmes informatiques.

Seule l'application de la génération automatique de texte entretient des liens plus étroits avec les approches fonctionnelles et formelles de la structure discursive. L'objectif principal en génération automatique de texte est de générer des énoncés qui permettent au lecteur/allocutaire de reconnaître les intentions du scripteur/locuteur de fournir un acte de langage spécifique. Ces approches visent à expliciter les relations entre les intentions du scripteur/locuteur et les actes de langage qui permettent de les réaliser.

La tâche de résumé automatique proposée par Marcu (2000) combine des techniques de surface et des méthodes statistiques.

Marcu (2000) propose d'aborder la tâche du résumé automatique de textes en tenant compte de la structure discursive de ceux-ci, suivant les travaux de la RST. Il démontre, d'une part, que la structure rhétorique des textes permet de mettre en lumière des changements de thématique dans le texte et, d'autre part, que celle-ci permet de déterminer la ou les phrases centrales en se basant sur la distinction opérée par la RST entre le noyau et le satellite d'une relation de discours, le noyau étant considéré comme un élément essentiel du texte et le satellite comme non essentiel et facultatif. Cette notion d'importance, appelée *nucléarité*, est utilisée par Marcu (2000) pour déterminer les segments les plus importants du texte, qui seront ainsi retenus pour l'obtention du résumé.

Les applications de TAL font appel à d'autres traitements automatiques discursifs tels que la résolution anaphorique. Le décalage entre l'apport des approches théoriques et le traitement en TAL pour cette tâche est important. Nous avons vu que les théories du discours ont beaucoup étudié les contraintes qui pèsent sur la résolution anaphorique avec la définition de la frontière droite. Les applications de TAL ne mettent pas en place de techniques d'analyse profonde des textes mais utilisent des méthodes de surface pour analyser les phrases et opèrent un calcul de saillance des unités référentielles basé sur la proximité de l'unité par rapport à l'élément anaphorique et la position syntaxique. Aucune approche de résolution anaphorique, à notre connaissance, ne fait appel à des structures de discours complexes.

1.4.2 Travaux récents sur la détection automatique des relations de discours

Étant donné que la structure du discours est basée sur des relations sémantiques et pragmatiques entre des unités de texte, détecter celle-ci automatiquement demeure un défi pour les recherches en TAL. Dans cette section, nous allons présenter quelques travaux récents ayant pour objectif d'atteindre automatiquement tout ou partie de ces structures discursives.

Généralement, l'approche adoptée consiste en un apprentissage automatique sur un corpus annoté en relations de discours selon diverses méthodes (approche probabiliste (Baldrige et Lascarides, 2005), approche d'apprentissage en logique du premier ordre (Subba et Di Eugenio, 2009)). Ces approches visent à détecter automatiquement tout (Baldrige et Lascarides, 2005; Subba et Di Eugenio, 2009) ou partie (les relations implicites (Lin *et al.*, 2009)) des structures discursives d'un texte. Pour ce faire, il est nécessaire d'employer des corpus enrichis automatiquement avec des informations lexicales, syntaxiques, sémantiques et manuellement avec des relations de discours. De tels corpus existent déjà pour l'anglais (le corpus DISCOR (Reese *et al.*, 2007), le PDTB (Carlson et Marcu, 2001)). Dans le cadre du projet ANNODIS, nous avons participé à la construction d'un tel corpus pour le français, nous permettant à l'issue de ce travail de thèse de bénéficier de données uniques pour le français. Nos objectifs seront moins ambitieux que ceux des travaux ci-dessus car nous souhaitons nous concentrer sur le marquage d'une relation de discours spécifique, la relation d'*Élaboration* et amorcer des études sur son repérage automatique.

Il est généralement admis que des informations issues de la structure discursive vont jouer un rôle important dans des applications de TAL. Par exemple, la reconnaissance des relations causales permet de répondre aux questions en «pourquoi» dans un système de question-réponse (Verberne *et al.*, 2006) ou les relations de *Contraste*, de *Résumé* et d'*Élaboration* sont utilisées dans des systèmes de résumé automatique ou des systèmes de détection des paraphrases. Leur détection automatique constitue donc un enjeu considérable.

1.5 Bilan et positionnement

Dans ce chapitre, nous avons présenté de nombreux aspects de l'analyse discursive. Dans notre thèse, nous adoptons une approche ascendante, qui part des plus petites unités du discours (UDE) pour atteindre la globalité du texte *via* des relations de discours. Nous souhaitons étudier plus précisément une relation de discours, la relation d'*Élaboration*. Les approches ascendantes que nous avons présentées ont travaillé sur la construction de modèle(s) visant à décrire, représenter, modéliser les discours dans leur globalité. Ils ont défini, décrit, modélisé les relations de discours nécessaires. Mais peu d'études ont porté sur l'analyse et la description approfondies d'une relation de discours. Nous pouvons citer à titre d'exemple les travaux sur la relation de *Contraste* (Busquets, 2007). Nous souhaitons, dans cette thèse, approfondir la description de la relation d'*Élaboration* en fournissant des analyses les plus fines possibles. Comme nous l'avons signalé en introduction, ce travail a pris sa source dans le cadre du projet ILF. La relation d'*Élaboration* avait été jusqu'alors peu étudiée. Le projet a fait émerger le peu de consensus existant sur les définitions de cette relation (Kleiber et Vassiliadiou, 2009), cette question sera développée dans le chapitre 2. Les projets d'annotation sur corpus des relations de discours ont permis de montrer que la relation d'*Élaboration* est d'une part très fréquente dans les textes (dans le corpus ANNODIS, 46% des relations annotées sont des relations d'élaboration) et d'autre part souvent confondues avec d'autres relations de discours (*Arrière-Plan*, *Explication*...). Ces deux constats confirment la nécessité d'approfondir la description et les définitions de cette relation, comme cela a été commencé dans les travaux de Prévot *et al.* (2009) dont nous ferons état dans le chapitre 3.

Parmi les approches ascendantes que nous avons présentées, nous avons opté pour une approche sémantique de la structure du discours, la Théorie des Représentations de Discours Structurées (SDRT). Plusieurs raisons ont guidé le choix de cette théorie.

D'abord, la SDRT propose un éventail d'une dizaine de relations de discours. Chacune des relations de discours est définie rigoureusement, comme nous l'avons présenté, avec des règles de déclenchement et des effets sémantiques. Une nouvelle relation de discours ne peut alors être intégrée dans le modèle qu'à condition que ses effets sémantiques soient différents de ceux des relations de discours déjà existantes. L'analyse avec le modèle SDRT repose sur un ensemble réduit mais également fermé de relations de discours, avec une analyse procédant par élimination. À priori, cela permet d'inférer dans tous les cas une relation de discours.

Ensuite, la SDRT est une théorie formelle qui vise à rendre compte de la manière dont on interprète un discours par le biais de sa représentation et de sa construction au moyen

de règles d'inférence rigoureuses. C'est à la fois un modèle explicatif qui cherche à expliciter la sémantique d'une séquence d'énoncés et un modèle prédictif qui cherche à fournir les moyens de dire qu'un énoncé sera valable ou non comme poursuite d'un discours.

Nous notons également que la SDRT accorde une place primordiale aux marquages des relations de discours en général et aux connecteurs en particulier. Les représentations logiques se construisent à partir de ce qui est disponible dans le texte (tandis que les intentions sont rarement signalées). Les règles d'inférence se basent sur différents types d'informations et sur la façon dont ces types d'informations interagissent entre eux. Cette approche sémantique implique de porter une attention extrêmement fine aux détails des réalisations linguistiques.

Enfin, le cadre de la SDRT nous paraît de plus très approprié pour notre travail car il s'agit d'une théorie en mouvement, pour laquelle de nombreux travaux sont produits visant à compléter, voire à améliorer, le modèle. Notre travail se situe donc dans une recherche dynamique.

Comme nous l'avons vu, une étape délicate de la construction d'un modèle de la cohérence consiste à établir une liste de relations de discours et leurs descriptions. Pour poursuivre cet état de l'art, nous allons présenter, dans les deux prochains chapitres, d'abord la notion d'élaboration, puis la relation d'*Élaboration*. Nous dégagerons des descriptions pour cette relation sur lesquelles nous nous appuierons ensuite dans nos analyses.

Chapitre 2

Les descriptions de l'élaboration

Sommaire

2.1	Historique de la notion d'élaboration	34
2.1.1	Grammaires traditionnelles	34
2.1.2	Prédicats rhétoriques	36
2.1.3	Analyse conversationnelle	39
2.1.4	Bilan	41
2.2	État de l'art des définitions données à l'Élaboration	41
2.2.1	Points communs et divergences	42
2.2.1.1	Apport d'informations supplémentaires	42
2.2.1.2	Identité ou différenciation?	44
2.2.2	Approches fonctionnelles	44
2.2.2.1	La relation d'Élaboration dans la Grammaire Fonctionnelle	44
2.2.2.2	La relation d'Élaboration dans la RST	50
2.2.2.3	La relation d'Élaboration chez Hovy et Maier (1991) . .	54
2.2.2.4	Bilan	56
2.2.3	Approches formelles	57
2.2.3.1	La relation d'Élaboration chez Hobbs (1979, 1985, 1990)	57
2.2.3.2	La relation d'Élaboration chez Kehler (2002, 2004) . . .	63
2.3	Élaboration en génération automatique de texte	65
2.4	Synthèse	67

Ce chapitre est d'abord consacré à la description de la notion d'élaboration. Le terme *élaboration* vient du verbe anglais *to elaborate*. Ce verbe a deux sens¹ :

- *elaborate*₁ est un verbe transitif qui se traduit en français par *élaborer*, il signifie *développer, construire*;
- *elaborate*₂ est un verbe intransitif qui se traduit en français par *donner des détails, entrer dans les détails*.

Le nom de la relation de discours est dérivé du second sens (*elaborate*₂) qui ne peut pas se traduire en français par *élaborer*, ce qui explique l'opacité sémantique du nom de la relation en français.

¹Dictionnaire *The New Penguin English* (2000).

Avant de parler de la relation de discours à proprement parler, nous allons remonter dans le temps (avant 1980) et nous intéresser aux prémices de la notion d'élaboration. Nous avons bénéficié du travail de collecte de relations (au sens large) de Hovy et Maier (1991) dans une trentaine de travaux de domaines divers (grammaire, compréhension et génération automatique de texte, analyse conversationnelle...). Concernant l'élaboration, ils ont relevé une vingtaine de travaux employant une ou des notions qui peuvent être rapprochées de celle-ci. Nous poursuivons ce travail pour la relation d'*Élaboration* avec l'ajout de références plus récentes. Les premiers travaux que nous présentons dans ce chapitre (Section 2.1) n'emploient pas le terme d'*élaboration* mais des notions proches de celle-ci si nous nous basons sur la définition de *elaborate*₂. Puis nous faisons un état de l'art (après 1980) des définitions qui ont été données à la relation de discours *Élaboration* identifiée comme telle (Section 2.2). Enfin, nous consacrons la dernière section (2.3) à la place de la relation d'*Élaboration* en génération automatique de texte. Nous souhaitons montrer, dans cette section, les choix retenus dans les définitions proposées dans le cadre d'une application spécifique (pour laquelle les travaux ont accordé un intérêt particulier aux relations de discours en général).

2.1 Historique de la notion d'élaboration

Les travaux présentés dans cette section ont deux points communs : ils sont anciens – ils datent d'avant 1980 – et ils n'emploient pas les termes de relation de discours, ni de relation d'*Élaboration*. Si nous souhaitons présenter ces travaux, c'est qu'ils sont précurseurs à la fois de la notion de relation de discours et de la relation d'*Élaboration*. Toutes les notions présentées dans les sections suivantes seront rapprochées de la notion d'élaboration sur la base du sens *elaborate*₂.

2.1.1 Grammaires traditionnelles

Nous commençons notre tour d'horizon par des travaux de grammairiens : Williams (1893) et Shipherd (1926) d'une part, puis Quirk et Greenbaum (1973) d'autre part. Les travaux de Williams et Shipherd concernent l'art de bien écrire des paragraphes. Les travaux de Quirk et Greenbaum concernent la construction de la phrase complexe, notamment au moyen de connecteurs.

Les travaux de Williams (1893) et Shipherd (1926) se situent dans la lignée des travaux d'Aristote. Ce dernier décrit les moyens mis à disposition du locuteur ou du scripteur pour construire des textes persuasifs. Par exemple, il distingue les *sylogismes* qui sont des types de raisonnement et les *exemples* qui fournissent des preuves à l'argumentation. Dans cette tradition, Williams (1893) et Shipherd (1926), grammairiens traditionnels, ont proposé d'assigner à chaque phrase d'un texte un prédicat afin de montrer aux apprenants comment bien construire des paragraphes.

Williams (1893) distingue plusieurs prédicats que nous allons rapprocher de la notion d'élaboration : *Topique*, *Amplification*, *General Illustration*, *Particular Illustration* et *Conclusion*. Malheureusement, nous n'avons pas accès aux définitions de ces prédicats mais seulement à une illustration de ceux-ci. Un prédicat classifiant est assigné à chaque phrase

du paragraphe, comme illustré dans l'exemple (1) (le prédicat est annoté à la fin de chaque phrase). Le prédicat attribué à chaque phrase correspond à une fonction que la phrase occupe au sein du paragraphe.

- (1) What, then, are the proper encouragements of genius? (*Topic*)
 I answer, subsistence and respect, for these are rewards congenial to nature. (*Amplification*)
 Every animal has an aliment suited to his constitution. (*General Illustration*)
 The heavy ox seeks nourishment from earth; the light chameleon has been supposed to exist on air. (*Particular Illustration*)
 (Williams, 1893) cité dans (McKeown, 1985, p. 24)
Quelles sont, alors, les véritables motivations des génies? (Topique) Je réponds, la subsistance et le respect, pour lesquels la nature offre une agréable récompense. (Amplification) Chaque animal a un mode d'alimentation qui convient à sa morphologie. (Illustration générale) Le boeuf, imposant, se nourrit de la terre; le caméléon, léger, est supposé exister dans l'air. (Illustration particulière) (Notre traduction)

La première phrase de (1) classifiée comme *Topique* donne le thème général abordé dans le paragraphe. Le topique est ici donné sous la forme d'une question. La suite du paragraphe, en répondant à la question, élabore le thème du paragraphe. La phrase suivante constitue une réponse directe à cette question, classifiée comme *Amplification*. L'*Amplification* est proche de la notion d'élaboration dans le sens où la phrase annotée *Amplification* donne une information plus spécifique en précisant quels sont les «véritables motivations des génies» données dans le *Topique*, c'est-à-dire «la subsistance et le respect». Puis, la troisième phrase est classifiée *Illustration générale*. Cette phrase est une illustration de la phrase précédente : «une des récompenses de la nature est de répondre aux besoins de subsistance des animaux en adaptant leur besoin en nourriture à leur constitution». Ensuite, la quatrième phrase est classifiée *Illustration particulière*. Cette phrase est également une illustration de la phrase précédente : «les bœufs étant lourds se nourrissent sur terre et les caméléons, qui sont légers, se nourrissent dans l'air». Nous considérons l'illustration comme un moyen d'entrer dans les détails, et donc d'élaborer. Enfin, nous souhaitons également rapprocher pour le moment le prédicat *Conclusion* qui n'est pas illustré ci-dessus de la notion d'élaboration, en tenant compte du fait qu'une conclusion est une façon de reformuler un contenu, en changeant de point de vue.

Williams (1893) propose une vision normative de l'écriture. Il décrit ce qu'il faut faire et ne pas faire dans l'écriture des paragraphes. Mais il ne dit rien sur comment combiner les phrases en paragraphes selon leur fonction. Williams (1893) cite des exemples de paragraphes bien écrits et annote les prédicats au sein de ces paragraphes, i.e. il fournit un typage de toutes les phrases du paragraphe bien écrit.

Cette approche vise à typer des phrases au sein d'un paragraphe et non pas à mettre ces phrases en relation comme dans les théories du discours que nous avons présentées dans le chapitre 1.

Dans le cadre d'une grammaire de l'anglais plus récente, Quirk et Greenbaum (1973) étudient la construction de la phrase complexe, i.e. les connections qui s'établissent entre deux propositions ou deux phrases, en identifiant la relation mise en place. Parmi les

configurations que nous allons rapprocher de la notion d'élaboration, se distinguent deux constructions tout à fait différentes. Les premiers cas que nous étudierons relèvent de la parataxe, plus exactement il s'agit des appositions non restrictives (ne permettant pas l'identification partielle ou totale du référent désigné par l'antécédent). La relation ici ne s'établit pas entre deux propositions ou phrases mais entre un nom et une apposition. Généralement l'apposition vient apporter des détails, permettant de nommer, désigner, identifier, reformuler, donner une description ou détailler une entité (Quirk et Greenbaum, 1973, p. 276-282) :

Appellation *The company commander, (that is to say) Captain Madison.*

Designation *Captain Madison, (that is to say) the company commander.*

Identification *A company commander, (namely) Captain Madison.*

Reformulation *He drew a triacontahedral, or thirty-sided figure.*

Attribution *The house, an imposing building dominating the street...*

Inclusion *Famous men (De Gaulle, Churchill, Roosevelt) have visited...*

Les seconds cas relèvent de la macro-syntaxe, c'est-à-dire de la façon dont on peut relier deux phrases ou deux propositions au moyen d'un connecteur (Quirk et Greenbaum, 1973, p. 288-291). Les connecteurs que nous listons ci-dessous concernent plusieurs facettes de l'élaboration : les connecteurs relevant de l'*Apposition* permettent de rentrer dans le détail en donnant un exemple ; les connecteurs relevant de la *Reformulation* permettent de reprendre le contenu de la phrase ou proposition précédente en d'autres termes ; les connecteurs relevant de la *Summation* ont également été retenus dans le sens où le résumé est un cas particulier de la reformulation (reformulation en un nombre de mots inférieurs) ; enfin nous retenons également les connecteurs relevant de l'*Énumération*, puisque nous avons l'intuition que lorsque l'on énumère, on détaille les éléments d'un tout. Nous reviendrons sur le lien entre structure énumérative et *Élaboration* dans le chapitre 6.

Apposition *for example, another way to putting it is..., an example would be...*

Reformulation/Remplacement *or rather, or in other words, a better way of putting it, it would be better to say, or again, or on the other hand, the alternative is, it might be better if...*

Summation *in brief, in all, I will sum up by saying, I shall conclude by saying...*

Enumeration *furthermore, first, finally, firstly, secondly, thirdly, above all, on top of it all, last but not least...*

Cette liste d'expressions nommées *connecteurs logiques* par Quirk et Greenbaum (1973) constitue bien évidemment un point de départ pour la recherche de marqueurs de la relation d'*Élaboration* que nous souhaitons mener dans cette thèse (voir sur ce point le chapitre 4 concernant la signalisation des relations de discours).

2.1.2 Prédicats rhétoriques

Les travaux de Grimes (1975) font partie de la tradition des «grammaires de cas» qui accordent à la sémantique une place plus importante qu'à la syntaxe dans la représentation

des phrases. Dans l'ouvrage *The Thread of Discourse*, Grimes (1975) étend l'analyse de la structure des phrases à l'analyse de la structure du discours.

Chez cet auteur, toutes les propositions dont les arguments ne sont pas reliés à leur prédicat via un rôle sémantique sont appelées des *propositions rhétoriques*. Les prédicats qui relient ces propositions sont appelés *prédicats rhétoriques*. Les prédicats rhétoriques sont récursifs : ils s'appliquent d'abord aux propositions, puis aux phrases, puis aux paragraphes jusqu'à couvrir de longues séquences de texte. Mais Grimes (1975) n'entre pas dans le détail de l'application des prédicats rhétoriques au-delà de la proposition et de la phrase. Il distingue trois familles de prédicats rhétoriques : *Supporting/Supplementary* pour les prédicats qui ajoutent des détails, expliquent ou justifient une information précédemment donnée ; *Setting* pour les prédicats qui situent un objet ou un événement dans l'espace et dans le temps ; et *Identification* pour les prédicats qui établissent et maintiennent la référence d'un objet.

Les prédicats rhétoriques tournés vers l'ajout de détails sont bien évidemment ceux qui nous intéressent, en lien avec la notion d'élaboration. Nous relevons les prédicats rhétoriques suivants (les prédicats relient les segments de texte (a) et (b)) :

Attributive «adds qualities or color to another predicate as center» (Grimes, 1975, p. 214) :

- (2) (a) Oh! but we was a tight-fisted hand at the grindstone, Scrooge! - (b) a squeezing, wrenching, grasping, clutching, covetous old sinner! Hard and sharp as flint, from which no steel had ever struck out generous fire; secret and self-contained, and solitary as an oyster. (C. Dickens, Christmas Carol).
 (a) Oh! il tenait bien le poing fermé sur la meule, le bonhomme Scrooge! - (b) Le vieux pêcheur branlant était un avare qui savait saisir fortement, arracher, tordre, pressurer, gratter, ne point lâcher surtout! Dur et tranchant comme une pierre à fusil dont jamais l'acier n'a fait jaillir une étincelle généreuse, secret, renfermé en lui-même et solitaire comme une huître. (Traduction Amédée Pichot)

Les propositions du segment (b) contribuent à ajouter des détails à la description du personnage *Scrooge*. Le prédicat *Attributive* correspond sémantiquement au type d'apposition *Attribution* de Quirk et Greenbaum (1973).

Equivalent «rather than adding information, simply restates it» (Grimes, 1975, p. 215) :

- (3) (a) we planned to leave on May 1, (b) the day of the spring celebrations.
 (a) Nous avons prévu de partir le 1er Mai, (b) le jour de la fête du printemps.
 (Notre traduction)

L'information subordonnée est présentée sous un aspect différent de l'information superordonnée. Le prédicat *Equivalent* est similaire au type d'apposition *Reformulation* chez Quirk et Greenbaum (1973).

Specifically «relates subordinate information that is semantically less inclusive to a center² that is more inclusive and therefore less precise» (Grimes, 1975, p. 215) :

²Élément superordonné.

- (4) (a) I heard a flock of birds flying south - (b) geese.
 (a) *J'ai entendu passer un vol d'oiseau qui allait vers le sud - (b) des oies.*
 (Notre traduction)

Le prédicat *Specifically* indique le passage d'une information plus générale «birds» vers une information plus spécifique «geese». La relation de spécificité s'applique ici entre les items lexicaux. Mais cette relation peut également s'appliquer entre des sous-arbres sémantiques plus larges, par exemple si l'on introduit une histoire en des termes très généraux, puis si l'on redit cette histoire en des termes plus spécifiques :

- (5) (a) Unkle George told me a story about a little girl and three bears. (b) It seems that there was this little girl named Goldilocks who lived in a house on the edge of the forest. One day... .
 (a) *L'oncle George m'a raconté une histoire au sujet d'une petite fille et de trois ours. (b) Il semble que cette petite fille appelée Goldilocks vivait dans une maison à la lisière de la forêt. Un jour...* (Notre traduction)

En anglais, des connecteurs tels que *namely* et *that is* permettent souvent d'introduire des sous-arbres subordonnés reliés à un sous-arbre dominant par le prédicat *Specifically* (Grimes, 1975). Ce prédicat semble correspondre au type d'apposition *Inclusion* de Quirk et Greenbaum (1973).

L'exemple suivant contient deux phrases reliées par le prédicat rhétorique *Specifically* :

- (6) He saved the day. He made three touchdowns.

La structure de cet exemple est représentée en Figure 2.1 avec l'ajout du prédicat *Specifically* comme un argument supplémentaire à l'état général «he saved the day» qui constitue le *center*. Cette notion de *center* indique que l'argument est superordonné et qu'il est relié à une autre proposition traitée comme un argument supplémentaire via un prédicat rhétorique :

- (a1 : he saved the day (omitting other details))
 (a2 : he made three touchdowns)
 (a3 : specifically a1 (center) a2)

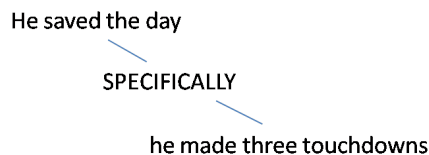


FIG. 2.1 – Structure de (6) chez Grimes (1975)

Les prédicats rhétoriques sont des précurseurs des relations de discours. Comme le souligne Grimes (1975), les prédicats rhétoriques, tout comme les relations de discours, permettent de relier entre elles des propositions mais également des portions de textes plus importantes via la récursivité. Les travaux de Reichman (1978) sur l'analyse conversation-

nelle, que nous allons présenter dans la section suivante, font justement intervenir la notion de relation sur des pans de textes plus larges que la proposition ou même la phrase.

2.1.3 Analyse conversationnelle

Dans ses travaux sur la cohérence des conversations, Reichman (1978) emploie également des relations pour lier entre elles des unités de texte. Plus exactement, ces relations s'établissent entre des *context spaces*. Le but des travaux de Reichman (1978) est de spécifier les étapes du processus par lequel les participants maintiennent la cohérence d'une conversation. C'est, à notre connaissance, le premier travail dans ce domaine qui fait appel à des relations afin de rendre compte de la cohérence des discours. Reichman décrit la conversation comme une séquence d'énoncés structurés hiérarchiquement. Ces séquences d'énoncés constituent les *context spaces* dont l'auteur distingue deux types : les *issue context spaces* qui sont des séquences d'énoncés décrivant des activités générales ou des états et les *event context spaces* qui sont des séquences d'énoncés décrivant des épisodes particuliers qui ont eu lieu dans le passé³.

Parmi les relations proposées par Reichman, deux peuvent être rapprochées de la notion d'élaboration : *Illustrative Relation* et *Restatement Relation*. Les relations sont définies par des schémas de réalisation. La relation *Illustrative* est une relation entre un *issue context space* (1.) et un ou plusieurs (caractère joker «+») *event context spaces* (2.) qui exemplifient l'*issue* précédente. Ce schéma est quelquefois (caractère joker «?») suivi de la relation *Restatement* par l'ajout d'un *issue context space* (3.) qui reformule le précédent *issue context spaces* :

1. Issue context space
2. (Event context spaces qui exemplifient (1)) +
3. (Issue context space qui reformule (1)) ?

L'exemple suivant illustre ce schéma avec les relations *Illustrative* et *Restatement* :

- (7) (a) **P** : I think a lot of progress has been made, not a lot - Progress has been made, even in my not being overly sensitive anymore to people's hangups, you know. If I see someone being totally insecure it's not my fault, you know, and I don't get that super involved in it
- (b) That - I'm actually referring to something last night. This girlfriend of mine, Debbie, was over and we were talking and I think I interrupted her, but I'm not sure, I'm really not sure if I did or not. But all of a - But anyway, all of a sudden her face took on all these funny - she shows her emotions - took on these funny features in a way. And I noticed it. And a couple of weeks ago, months ago, I would've felt very guilty. What did I wrong, etc. etc. And this time I didn't (...) Um, I was aware of it but I didn't start feeling guilty myself
- (c) You know like with - when I went to the movies with Cindy and Bob and I felt she was ignoring him.(...)
- (d) I mean I noticed it but I didn't start feeling guilty about it
- (Reichman, 1978, p. 297-298)

³Ce qui nous renvoie respectivement aux notions d'état et d'événement.

- (a) **P** : Je pense que beaucoup de progrès ont été faits, non pas beaucoup - Des progrès ont été faits, même concernant ma sensibilité extrême aux complexes des gens, vous savez. Si je vois quelqu'un qui manque totalement d'assurance ce n'est pas ma faute, vous savez, et je ne suis plus tant super impliquée que ça
- (b) ça - je fais actuellement référence à un événement de la nuit dernière. Cette amie, Debbie, était découragée et nous étions en train de parler et je pense que je l'ai interrompue, mais je ne suis pas sûre, je ne suis vraiment pas sûre si je l'ai fait ou pas. Mais tout d'un - Mais en tout cas tout d'un coup son visage a pris une expression bizarre - elle montre ses émotions - pris cette expression bizarre en fait. Et je l'ai remarqué. Et il y a quelques semaines, quelques mois plus tôt, j'aurais ressenti beaucoup de culpabilité. Qu'ai-je fait de mal, etc. etc. Et cette fois je ne l'ai pas fait (...) Hum, j'étais consciente de ça mais je n'ai pas commencé à me sentir coupable
- (c) Vous savez comme quand avec - quand je rentrais du cinéma avec Cindy et Bob et que j'ai ressenti qu'elle était en train de l'ignorer. (...)
- (d) Je veux dire je l'ai remarqué mais je n'ai pas commencé à me sentir coupable à ce sujet. (Notre traduction)

La relation *Illustrative* s'applique entre les segments (a) et (b) et entre les segments (a) et (c). Reichman (1978) distingue deux types d'illustration : *Reference Illustrative* quand l'*event space context* est déjà connu de l'allocutaire (a-c) et *Full Illustrative* lorsque l'allocutaire n'a pas de connaissances préalables (a-b). Enfin, la relation *Restatement* s'applique entre les segments (a) et (d).

Les objectifs de Reichman (1978) sont de mettre au jour les règles et les régularités qui gouvernent les séquences d'énoncés dans une conversation. Il observe que des manifestations linguistiques accompagnent souvent les transitions entre *context spaces*, par exemple les marqueurs du discours (qu'il appelle *clue words*) comme *like* dans l'exemple suivant :

- (8) (a) **F** : I'm starting to, you know, to get more insight about dreams. And they're so revealing about where you're at.
- (b) They really are. **Like** Susan and I spent an hour the other night just dissecting a dream. **C** : I know. I should write them down. **F** : And she wrote it down, and when I read it I saw new things into it. And it really reflected so much of where she's at in therapy and everything.
- (a) **F** : J'ai commencé, vous savez, à être plus éclairé au sujet de mes rêves. Et ils étaient si révélateurs à propos d'où tu en es
- (b) Ils le sont vraiment. **Comme** lorsque Susan et moi avons passé une heure l'autre nuit juste à disséquer un rêve. **C** : Je sais. Je devrais les écrire. **F** : Et elle les a écrits, et quand je les lis je vois de nouvelles choses. Et ils reflètent tellement où elle en est de la thérapie et tout le reste. (Reichman, 1978, p. 308)

La relation qui s'établit entre les deux *context spaces* (a) et (b) de l'exemple ci-dessus est *Illustrative Relation*. Elle est signalée par le marqueur du discours *like*.

2.1.4 Bilan

Ce premier aperçu autour de la notion d'élaboration est riche d'enseignement.

D'abord, nous notons que la notion d'élaboration apparaît dans de très nombreux travaux sur le discours. Nous notons également que l'inventaire que nous venons de faire montre la variété des relations sémantiques que nous rapprochons de l'élaboration, comme détailler une entité, l'illustrer, l'exemplifier, la reformuler... Parmi ces différentes relations, les études que nous avons présentées montrent que soit un événement, soit une entité (les appositions de Quirk et Greenbaum (1973)) peut être élaboré dans le discours. Nous reviendrons sur cette distinction tout au long de ce chapitre avant de traiter la question plus en profondeur dans le chapitre suivant.

Ce premier aperçu montre également que ce type de relation ne lie pas uniquement des propositions et des phrases mais aussi des séquences de propositions ou de phrases (Reichman, 1978; Grimes, 1975). Ensuite, nous retenons de ces approches que la notion de relation est fortement liée à la notion de marqueurs lexicaux avec la grammaire de Quirk et Greenbaum (1973) qui part du connecteur pour préciser la relation que celui-ci exprime. Grimes (1975) et Reichman (1978) font également allusion aux marqueurs du discours. Nous reviendrons sur la question de la signalisation des relations de discours dans le chapitre 4.

La prochaine section va concerner les définitions qui ont été données à la relation d'*Élaboration* dans les théories dont il a été question au chapitre 1. Les premiers à avoir introduit la notion de relation de discours, appelée relation de cohérence, sont Hirst (1981) et Hobbs (1979). Les relations de cohérence sont utilisées pour une tâche précise de résolution des anaphores. Hirst (1981) et Hobbs (1979) sont également les premiers à employer la relation d'*Élaboration*.

2.2 État de l'art des définitions données à l'*Élaboration*

Comme nous l'avons vu au chapitre 1, les théories du discours s'accordent sur le fait qu'il existe des relations de discours qui s'établissent entre des segments de discours. Cependant, ces différentes théories ne s'accordent pas toujours sur le nombre de relations nécessaires pour rendre compte de la structure d'un discours, allant de deux chez Grosz et Sidner (1986) à une hiérarchie d'approximativement soixante-dix relations (organisées dans une hiérarchie de spécificité croissante) chez Hovy et Maier (1991), ni sur leur classification, dénomination et définition. La relation d'*Élaboration* n'échappe pas au manque de consensus général qui existe sur les relations de discours. Ces divergences rendent extrêmement difficile la tâche de repérage de celles-ci dans les textes. Ceux qui s'aventurent dans cette entreprise périlleuse (nous pouvons citer le projet ANNODIS (Péry-Woodley *et al.*, 2009), les travaux du Penn Discourse Tree Bank (Prasad *et al.*, 2007) et le projet DISCOR (Reese *et al.*, 2007) entre autres...) doivent avant tout faire un travail de mise au point sur le nombre de relations de discours nécessaires et les définir avec rigueur.

Nous proposons dans ce chapitre d'examiner en détail les définitions qui ont été données à la relation d'*Élaboration* dans plusieurs théories du discours⁴ (la Grammaire Fonctionnelle

⁴Ce chapitre doit beaucoup au projet ILF «Relations de cohérence et fonctionnement des anaphores»

(Halliday, 1985, entre autres), la RST (Mann et Thompson, 1987, entre autres), la synthèse de Hovy et Maier (1991), les théories de Hobbs (1990, entre autres) et Kehler (2002)). Nous réservons les définitions et/ou descriptions de la relation dans la SDRT au chapitre 3. Nous avons vu dans la section 2.1 que la notion d'élaboration peut être rapprochée de plusieurs autres notions, telles que l'exemplification et la reformulation. Certaines théories vont faire le choix de distinguer ces notions en proposant plusieurs relations de discours (la RST (Mann et Thompson, 1987), les théories de Hobbs (1990) et Kehler (2002)) tandis que d'autres vont faire le choix d'une seule relation de discours qui englobe toutes les notions proches de l'élaboration ((Halliday, 1985) et la SDRT (Asher, 1993; Asher et Lascarides, 2003)). Une troisième possibilité est d'organiser les relations dans une hiérarchie de spécificité croissante comme proposé par Hovy et Maier (1991).

Ce travail nous paraît d'autant plus nécessaire sur ladite relation d'*Élaboration*, que celle-ci est très fréquemment annotée dans les textes (presque 50% des annotations dans le projet ANNODIS). Les projets d'annotation que nous avons mentionnés ci-dessus ont également signalé que les taux d'accord sur l'annotation de la relation d'*Élaboration* sont parmi les plus mauvais.

Nous espérons que cet exposé des définitions données à la relation d'*Élaboration* permettra d'éclairer la relation au sein de chacune des théories qui seront présentées mais aussi d'éclairer la relation d'*Élaboration* elle-même. Nous nous demandons cependant s'il est pertinent de parler de la relation d'*Élaboration* en elle-même car une relation de discours prend son sens au sein d'une théorie où celle-ci se trouve naturellement délimitée par d'autres relations de discours. Donc, le but de ce chapitre sera plutôt de donner une vision de tout ce qui, dans le discours, peut se trouver sous l'étendard de l'*Élaboration*. Mais avant de nous intéresser aux théories du discours et aux définitions de la relation d'*Élaboration*, nous allons proposer une synthèse des points communs et des divergences majeures que nous avons rencontrés dans les différentes définitions de la relation que nous avons analysées.

2.2.1 Points communs et divergences

2.2.1.1 Apport d'informations supplémentaires

La relation d'*Élaboration* pâtit d'un manque de consensus général au sein des théories du discours. Toutefois, l'examen des quelques théories que nous avons envisagées montre qu'il existe au moins un consensus, pour les cas typiques, sur le fait que la relation s'établit entre deux segments de discours dont le second donne des informations supplémentaires sur le premier. Les théories du discours s'accordent également sur le fait que le premier segment apporte l'information essentielle et que le deuxième segment apporte des détails, précisions ou spécifications sur le premier segment. Dans les termes de la RST (Mann et Thompson, 1987), le segment élaboré est le noyau et le segment élaborant est le satellite. Les nombreuses confusions que nous avons évoquées portent essentiellement sur la façon dont on définit les informations supplémentaires et sur l'identification de cet élément auquel on apporte des informations supplémentaires. Les exemples suivants vont illustrer deux types de confusion possibles :

dirigé par F. Cornish et au volume 19-2 de JFLS (2009) qui regroupe les travaux des membres du projet.

- (9) (a) Paul s'est promené dans le parc. (b) Il n'arrivait plus à travailler.
- (10) (a) Paul s'est promené dans le parc. (b) Le soleil commençait à se coucher au bord du lac.
- (11) (a) Paul s'est promené dans le parc. (b) Il a traversé le petit bois jusqu'à la rivière.

En adoptant le point de vue d'un lecteur lors de l'interprétation de ces discours, nous reconnaissons des relations qui s'établissent entre les segments (a) et (b). Nous inférons ces relations en nous basant sur des connaissances du monde, des connaissances en sémantique lexicale et grammaticale et non sur la base de marqueurs du discours. Pour être plus précis, nous pouvons établir un rapport entre ce qui est dit dans le segment (a) et ce qui est dit dans le segment (b) ; ce qui assure la cohérence ou la pertinence de l'ensemble.

On pourrait dire pour chacun des exemples ci-dessus que le segment (b) apporte des informations supplémentaires sur le segment (a). Dans l'exemple (9), le segment (b) donne des informations supplémentaires en explicitant les raisons de la promenade. Dans (10), le segment (b) donne des informations supplémentaires en donnant des indications sur les circonstances de la promenade. Enfin, dans (11), le segment (b) donne des informations supplémentaires en donnant des détails sur une étape ou sur la totalité (redescription de l'événement) de la promenade. Nous voyons, avec ces trois exemples, trois façons très différentes de comprendre la notion trop vague d'« informations supplémentaires ». Cependant, dans les exemples (9), (10) et (11), il nous semble difficile de parler d'*Élaboration*. Dans de nombreuses théories du discours, il existe effectivement des relations différentes pour analyser ces exemples. Pour des cas comme (9), lorsqu'un segment (b) introduit une ou des causes possible(s) du segment (a), la relation d'*Explication* est établie. Pour des cas comme (10), lorsqu'un segment (b) précise les circonstances dans lesquelles s'est déroulé l'événement décrit dans le segment (a), la relation d'*Arrière-plan* est établie. Enfin, pour des cas comme (11), lorsqu'un segment (b) donne une étape ou une redescription de l'événement décrit dans le segment (a), la relation d'*Élaboration* est établie. Mais nous verrons qu'il existe réellement des confusions entre la relation d'*Élaboration* et respectivement les relations d'*Explication* et d'*Arrière-plan* lorsque l'on cherche à annoter ces relations dans des textes attestés⁵.

Ces trois exemples (9), (10) et (11) seront réutilisés dans la suite de ce chapitre pour tester les définitions données dans les différentes théories que nous allons présenter. Nous souhaitons également confronter ces théories à l'exemple suivant adapté d'un exemple de Kamp et Rohrer (1983) par Bras (2007) :

- (12) (a) Mixel a escaladé le Vignemale hier matin. (b) Il est parti du refuge des Oulettes au lever du jour, (c) puis il a passé la Hourquette d'Ossoue vers 9h, (d) et il est arrivé au sommet vers midi. (Bras, 2007, p. 232)

Dans cet exemple, les segments (b), (c) et (d) décrivent des étapes de l'événement décrit dans le segment (a). Ce type d'exemple est souvent cité comme étant un cas typique d'*Élaboration*. Nous souhaitons donc analyser cet exemple au moyen des définitions fournies dans les théories que nous présenterons dans la suite de ce chapitre.

⁵ Cf. chapitre 9 sur la relation d'*Élaboration* dans le projet ANNODIS.

2.2.1.2 Identité ou différenciation ?

Il existe un consensus sur le fait que, dans les cas d'*Élaboration*, il y a un apport informationnel et que cet apport est conçu comme étant mineur. L'information essentielle est donnée dans le premier segment. Le deuxième segment apporte des détails, des précisions ou des spécifications sur le premier segment. En fait, avec le second segment, on ne sort pas du cadre décrit dans le premier segment :

On reste donc dans le «même», mais avec une phrase «autre» : la seconde phrase ou phrase élaborante est à la fois identique à et différente de la première ou phrase élaborée. (Kleiber et Vassiliadiou, 2009, p. 185)

Les définitions de la relation d'*Élaboration* présentent le segment élaborant⁶ comme étant à la fois identique et différent du segment élaboré. Identique dans le sens où le segment élaborant reste dans le cadre de la situation décrite par le segment élaboré et différent dans le sens où le segment élaborant apporte des détails, des précisions ou des spécifications sur le segment élaboré.

Les divergences entre les théories sont présentées par Kleiber et Vassiliadiou (2009) comme une tension entre le pôle de l'identité et le pôle de la différence. Effectivement les définitions de la relation d'*Élaboration* dans les différentes théories vont soit privilégier l'identité entre les deux segments, comme par exemple dans (Hobbs, 1990), soit mettre l'accent sur l'information supplémentaire apportée par le deuxième segment, comme par exemple dans (Mann et Thompson, 1987).

À la lumière des notions d'identité et de différence, nous allons présenter dans les sections suivantes les définitions données à la relation dans la Grammaire Fonctionnelle de Halliday (1985), dans la RST (Mann et Thompson, 1987), chez Hovy et Maier (1991), chez Hobbs (1990) et chez Kehler (2002), puis dans la SDRT (Asher, 1993; Asher et Lascarides, 2003) dans le chapitre 3.

2.2.2 Approches fonctionnelles

Dans cette section, nous focaliserons notre attention sur la définition de la relation de discours *Élaboration* dans trois théories du discours fonctionnelles, la Grammaire Fonctionnelle (Halliday, 1985, 1994; Thompson, 2004), la RST (Mann et Thompson, 1986, 1987, 1988), puis chez Hovy et Maier (1991). Nous verrons dans cette section que la définition des relations de discours dans ces théories laisse une part majeure à l'intuition et à l'interprétation des lecteurs.

2.2.2.1 La relation d'*Élaboration* dans la Grammaire Fonctionnelle

Dans la Grammaire Fonctionnelle (Halliday, 1985, 1994; Thompson, 2004), les unités d'analyse sont les propositions complexes. Etant donné une proposition, celle-ci peut prendre part à une construction avec une autre proposition, les deux formant alors une proposition complexe. Les auteurs suggèrent trois façons de construire une proposition complexe à partir d'une proposition de départ : l'élaborer (*elaborate*), l'étendre (*expand*) ou l'augmenter (*enhance*) avec des informations sur le temps, les raisons, les conditions,

⁶Nous parlons de segments (qui peuvent être des propositions, des phrases, des successions de phrases) tandis que Kleiber et Vassiliadiou (2009) parlent uniquement de phrases.

etc. La définition d'*Élaboration* présente les deux aspects d'identité et de différence sans les distinguer explicitement :

In Elaboration, one clause elaborates on the meaning of another by further specifying or describing it. The secondary clause does not introduce a new element into the picture but rather provides a further characterization on one that is already there, restating it, clarifying it, refining it, or adding a descriptive attribute or comment. The thing that is elaborated may be the primary clause as a whole, or it may be just some part of it - one or more of its constituents. (Halliday, 1985, p. 203)

An elaborating clause does not add any essentially new element to the message but gives more information about what is already there. It may relate to the whole message, or just one part of the message ; and it may restate it ; or it may clarify or exemplify it ; or it may add extra information about its attributes, including the speaker's comment. (Thompson, 2004, p. 204)

Les deux définitions s'accordent sur le fait que l'*Élaboration* peut porter sur le segment entier ou sur une partie de celui-ci⁷. Ce point sera plus approfondi dans la section 3.3.1 du chapitre 3 suivant où sera détaillée la distinction entre les élaborations d'éventualité et les élaborations d'entité.

Ces définitions présentent le contenu du segment élaborant comme étant à la fois identique et différent du segment élaboré. L'identité entre le contenu des deux segments est exprimée dans la première citation par «*The secondary clause does not introduce a new element into the picture*» et dans la deuxième par «*An elaborating clause does not add any essentially new element to the message*». La différence est exprimée par «*... but rather provides a further characterization on one that is already there*» dans la première et par «*but gives more information about what is already there*» dans la deuxième.

Les deux définitions s'accordent sur le fait que le deuxième segment n'apporte pas d'éléments nouveaux, sans préciser exactement la nature de ces éléments nouveaux. La seconde partie de la définition nous paraît plus éclairante : les détails apportés par le segment élaborant portent sur un élément déjà présent dans le segment élaboré.

Les deux définitions proposent ensuite une liste de «façons» d'élaborer dans le discours : reformuler, clarifier, raffiner, décrire, exemplifier, commenter. On retrouve dans cette définition les notions proches de l'élaboration que nous avons présentées dans la section 2.1 sur l'historique de la notion d'élaboration.

Nous proposons d'appliquer ces définitions aux trois exemples que nous avons introduits dans la section 2.2.1 :

- (9) (a) Paul s'est promené dans le parc. (b) Il n'arrivait plus à travailler.
- (10) (a) Paul s'est promené dans le parc. (b) Le soleil commençait à se coucher au bord du lac.
- (11) (a) Paul s'est promené dans le parc. (b) Il a traversé le petit bois jusqu'à la rivière.

⁷Nous notons de plus que la première citation parle de la proposition : «*The thing that is elaborated may be the primary clause as a whole, or it may be just some part of it - one or more of its constituents.*» tandis que la seconde parle du message : «*It may relate to the whole message, or just one part of the message.*».

À partir des définitions de Halliday (1985) et Thompson (2004), nous pouvons reconnaître *Élaboration* dans chacun des cas ci-dessus. Le segment (b) dans (9) donne une description ou un commentaire sur «Paul». Le segment (b) dans (10) donne une description du «parc» au moment de la promenade. Enfin dans (11), le segment (b) est une clarification de l'événement décrit dans le segment (a). Or, comme nous l'avons montré dans la section 2.2.1, cette analyse n'est pas la plus convaincante étant donné que d'autres relations de discours peuvent préciser les liens qui s'établissent entre les segments des deux premiers exemples (*Explication* pour (9) et *Arrière-plan* pour (10) dans de nombreuses théories du discours) et dans le cadre de la Grammaire Fonctionnelle, la relation *Enhancement* peut être vérifiée dans (9) et (10) pour lesquels le deuxième segment spécifie des aspects de la proposition dominante tels que le temps, les raisons (9), les conditions (10)...

Il nous semble que ces deux définitions manquent de clarté : d'abord, il faut préciser quels sont ces éléments nouveaux et/ou essentiels qui ne peuvent pas être contenus dans le segment élaborant ; ensuite, il faut préciser les éléments du segment élaboré qui peuvent faire l'objet d'une élaboration.

Concernant l'exemple (12) dans lequel les segments élaborants donnent des étapes de l'événement décrit dans le segment élaboré (a), nous pouvons dire que les segments (b), (c) et (d) viennent *clarifier* ou *affiner* le message contenu dans le segment (a) :

(12) (a) Mixel a escaladé le Vignemale hier matin. (b) Il est parti du refuge des Oulettes au lever du jour, (c) puis il a passé la Hourquette d'Ossoue vers 9h, (d) et il est arrivé au sommet vers midi. (Bras, 2007, p. 232)

Pour approfondir notre compréhension de ces définitions, nous proposons de regarder les exemples d'*Élaboration* qu'Halliday (1985) et Thompson (2004) analysent. Aucun de ces exemples n'est similaire aux exemples (11) (redescription d'un événement) et (12) (étape d'un événement). De plus, nous allons voir que les exemples qu'ils proposent confirment notre première intuition sur les définitions. Ces derniers ne correspondent pas toujours à l'idée que nous nous faisons de l'élaboration.

Halliday (1985) et Thompson (2004) distinguent deux constructions des propositions complexes : la parataxe (coordination et juxtaposition de propositions) et l'hypotaxe (subordination de propositions). La relation d'*Élaboration* peut relier deux propositions simples (a) et (b) en une proposition complexe de façon paratactique ou hypotactique⁸.

Parataxe Du côté des propositions élaborantes paratactiques, Halliday (1985) et Thompson (2004) présentent des cas de juxtaposition.

Halliday (1985) distingue trois types d'*Élaboration* et donne des marqueurs qui peuvent expliciter la relation entre les deux propositions. Ces marqueurs peuvent apporter plus de cohésion mais généralement les deux propositions sont simplement juxtaposées. De ce fait, Halliday (1985) fait remarquer qu'il est parfois difficile de décider si les deux propositions juxtaposées forment une proposition complexe ou non. Toutefois, si la relation d'*Élaboration* est présente entre les deux propositions, cela constitue un indice pour les regrouper dans

⁸La proposition (a) est la proposition élaborée et la proposition (b) est la proposition élaborante.

un segment complexe. Le tableau 2.1 récapitule les définitions, exemples et marqueurs des trois types de la relation d'*Élaboration* dégagés par Halliday (1985).

Définition	Exemples	Marqueurs
Exposition <i>La seconde proposition reformule la thèse de la première en d'autres termes pour la présenter d'un point de vue différent ou pour renforcer le message.</i>	(I) (a) Each argument was fatal to the other : (b) both could not be true. / (II) (a) The clock doesn't go ; (b) it's not working.	in other words, that is to say, or (rather), i.e.
Exemplification <i>La seconde proposition reformule la thèse de la première en devenant plus spécifique ou en donnant un exemple.</i>	(III) (a) Your face is the same as everybody else has - (b) the two eyes so, nose in the middle, mouth under. / (IV) (a) We used to have races - (b) we used to have relays	for instance, for example, in particular, e.g.
Clarification <i>La seconde proposition clarifie la thèse de la première en confirmant celle-ci au moyen d'explication ou de commentaire explicatif.</i>	(V) (a) Alice could only look puzzled : (b) she was thinking of the pudding. / (VI) (a) I wasn't surprised - (b) it was what I had expected.	in fact, actually, indeed, at least, viz.

TAB. 2.1 – Exemples de Halliday (1985, p. 203)

Parmi les exemples donnés dans le tableau 2.1, les deux premiers groupes (exposition et exemplification) nous semblent effectivement correspondre à l'idée que nous nous faisons de la relation d'*Élaboration*. Nous notons que deux propositions uniquement juxtaposées sont reliées implicitement par la relation pour former une proposition complexe. Les marqueurs de la troisième colonne permettent, quant à eux, d'explicitement la relation entre les deux propositions.

Concernant le deuxième groupe, Halliday et Mathiessen (2004) précisent que la relation d'*Exemplification* est souvent accompagnée d'une relation lexicale, généralement une relation d'hyponymie ou de méronymie comme dans (IV) entre «races» et «relays» ou dans (III) entre «face» et «eyes», «nose», «mouth». La présence d'une relation lexicale d'hyponymie ou de méronymie entre les items des deux propositions a certainement amené Halliday (1985) et Halliday et Mathiessen (2004) à considérer ce dernier exemple (III) comme relevant de la relation d'*Exemplification*. Pourtant, nous notons qu'aucun marqueur de la relation ne peut s'y insérer tandis que les marqueurs de la relation d'*Exposition* semblent plus appropriés pour expliciter la relation entre les deux propositions.

En ce qui concerne le troisième groupe (*Clarification*), il s'éloigne, de notre point de vue, de la notion d'*élaboration*. Dans les exemples donnés ci-dessus, la deuxième proposition vient expliquer ou donner des preuves du contenu de la première proposition. Le marqueur

because prototypique de la relation d'*Explication* peut être inséré entre les propositions de chacun des exemples au même titre que les autres marqueurs signalés dans la troisième colonne.

Hypotaxe Du côté des propositions élaborantes hypotactiques, Halliday (1985) et Thompson (2004) évoquent les propositions relatives non restrictives⁹ qui ajoutent des informations supplémentaires à propos d'un élément du message et les propositions non finies (partiales (-ing), propositions infinitives...).

Halliday (1985) décrit la fonction des relatives non restrictives comme donnant un commentaire descriptif de la première proposition :

- (13) (a) They decided to cancel the show, (b) which upset everybody alike. (Halliday, 1985, p. 204)
 (a) *Ils ont décidé d'annuler le spectacle, (b) ce qui a vraiment dérangé tout le monde.*

Dans cet exemple, nous nous éloignons de la notion d'élaboration, la deuxième proposition étant plutôt comprise comme un résultat de l'événement décrit dans la première proposition.

L'antécédent peut être la proposition entière comme (13) et les deux exemples suivants :

- (14) (a) If I ever did fall off - (b) which there's no chance of... (Halliday, 1985, p. 204)
 (a) *Si jamais j'étais tombé - (b) ce qui n'a aucune chance de se produire...*
- (15) (a) From then on we started winning prizes, (b) which turned out to be very easy. (Halliday, 1985, p. 204)
 (a) *À partir de ce moment là, on a commencé à gagné des prix, (b) ce qui s'est avéré très facile.*

Dans ces deux exemples, le pronom relatif reprend le contenu de la première proposition. Cependant, le contenu de la proposition relative ne nous semble pas élaborer le contenu de la première proposition mais plutôt fournir un commentaire. Les commentaires sont inclus dans la définition d'*Élaboration* dans le cadre de la Grammaire Fonctionnelle. Pourtant, nous ne souhaitons pas les considérer comme des *Élaborations* car nous pensons que les commentaires sont des évaluations produites par un personnage, un scripteur ou un locuteur sur un message précédent. Dans (14), le locuteur évalue la condition «Si jamais je tombe» comme ayant une probabilité zéro et dans (15), le locuteur évalue le fait que «gagner des prix» s'est avéré très facile.

L'antécédent peut également être un groupe nominal :

- (16) (a) She was hard at work on the white kitten, (b) which was lying quite still. (Halliday, 1985, p. 205)
Elle travaillait dur sur le chaton blanc, (b) qui était allongé tout tranquillement.

⁹Les relatives non restrictives ne permettent pas l'identification (partielle ou totale) du référent désigné par l'antécédent, contrairement aux relatives restrictives, comme dans l'exemple suivant : *La rue qui longe la mairie est en travaux.*

- (17) (a) I was further upset by his voice, (b) which was loud, harsh and hoarse. (Thompson, 2004, p. 205)
 (a) *J'étais encore plus agacé par sa voix, (b) qui était forte, sévère et rauque.*

Dans ces deux exemples, la proposition relative fournit effectivement une caractérisation d'un objet du discours, respectivement «the white kitten» et «his voice». Les deux propositions sont donc, dans les deux cas, reliées par la relation d'*Élaboration* selon la définition de la relation. Cependant, de notre point de vue, la proposition relative dans (16) donne plutôt un arrière-plan à l'événement décrit dans la première proposition et la proposition relative dans (17) donne plutôt une explication à l'état décrit dans la première proposition.

Halliday (1985) fait remarquer que toutes les propositions relatives ne sont pas reliées à la proposition principale au moyen de l'*Élaboration*. Nous reviendrons sur ce sujet lors de la présentation dans le chapitre suivant (section 3.3.1) de la relation d'*Élaboration d'entité* qui relie souvent une proposition relative à une entité de la proposition principale.

Concernant les propositions non finies, nous sommes d'accord avec l'analyse de Halliday (1985) :

- (18) (a) I worked for a local firm at that time, (b) selling office equipment. (Halliday, 1985, p. 206)
 (a) *Je travaillais pour une entreprise locale à ce moment là, (b) vendant des équipements de bureau.*
- (19) (a) There was a real fire there, (b) blazing away just as brightly. (Halliday, 1985, p. 206)
 (a) *Il y avait un véritable feu à cet endroit, (b) brillant d'une lueur vive.*

La proposition non finie dans (18) et (19) vient élaborer la proposition principale, en précisant le métier pour le premier exemple et l'intensité du feu pour le second.

Au sujet des propositions non finies, Thompson (2004) précise qu'il est souvent difficile de distinguer si la proposition élabore (*elaborate*), donne des informations sur les circonstances, les raisons, les conditions (*enhance*) ou prolonge (*extend*) la proposition précédente :

- (20) (a) I was scared of the changes, (b) not knowing what life would be like (Thompson, 2004, p. 205)
J'étais effrayé par les changements, ne sachant pas de quoi la vie serait faite

Effectivement, si la deuxième proposition spécifie en quoi le locuteur est craintif; dans ce cas-là, nous avons une *Élaboration*. S'il indique pourquoi le locuteur est craintif; dans ce cas, nous avons une *Explication* (*'enhancing'*). Enfin, s'il exprime une autre réaction en plus d'être craintif; alors nous avons une *Extension* (*'extending'*), autrement dit une proposition qui prolonge la proposition précédente¹⁰.

Dans le cadre de la Grammaire Fonctionnelle, les auteurs cherchent à déterminer les relations qui permettent de construire une proposition complexe à partir de deux propositions

¹⁰En français, seules les deux premières interprétations semblent possibles à moins d'avoir la construction *tout en + gérondif* : «J'étais effrayé par les changements, tout en ne sachant pas de quoi la vie serait faite», ce qui exclut alors les deux autres interprétations.

simples. Ils explorent donc les constructions syntaxiques pour déterminer des corrélations entre constructions syntaxiques et relations sémantiques et discursives. D'après leurs travaux, nous pouvons conclure que deux propositions reliées par la relation d'*Élaboration* seront dans une structure syntaxique particulière (juxtaposition, relative non restrictive, proposition en *-ing* et infinitive). Cependant, nous notons que si deux propositions sont dans cette structure syntaxique particulière, elles ne seront pas forcément reliées par la relation d'*Élaboration*.

Sans toutefois accorder une importance maximale aux marqueurs du discours, nous trouvons dans ce travail l'idée qu'une relation peut être explicitée par une configuration linguistique particulière. Nous présenterons l'importance accordée aux marqueurs du discours et autres indices linguistiques de la signalisation discursive dans les travaux sur le discours dans le chapitre 4.

Les travaux présentés dans cette section constituent évidemment un point de départ pour ce travail de thèse étant donné que nous nous sommes donnée pour objectif de déterminer les moyens mis en oeuvre par les scripteurs pour exprimer la relation d'*Élaboration* et les moyens mis en oeuvre par les lecteurs pour la reconnaître. Nous examinerons les structures syntaxiques et les marqueurs lexicaux, mais aussi les relations lexicales (qui ont été évoquées dans cette section pour la sous-relation d'*Exemplification*) et les structures textuelles (telles que les structures énumératives).

Alors que les travaux de la Grammaire Fonctionnelle accordent une part importante au rôle des éléments constitutifs du texte (structure syntaxique, marqueurs lexicaux...) dans l'organisation des messages, la RST prend ses distances avec les éléments qui relèvent de la forme linguistique. Elle tient compte dans ses définitions des intentions présumées du scripteur et des effets que celui-ci entend produire sur le lecteur. Elle accentue la place importante de l'intuition et de l'interprétation que l'on trouve déjà dans la Grammaire Fonctionnelle.

2.2.2.2 La relation d'*Élaboration* dans la RST

La définition de la relation d'*Élaboration* dans la RST, comme toutes les définitions dans cette théorie, pose des contraintes sur le noyau (N), le satellite (S) et l'ensemble noyau + satellite (N+S) mis en relation, puis précise les effets de la relation et le locus de cet effet :

Constraints on N : None

Constraints on S : None

Constraints on the N+S combination : S presents additional detail about the situation or some element of the subject matter which is presented in N, or is inferentially accessible from N, in one or more of the ways listed below. In the list, if N presents the first member of any pair, then S includes the second :

- set - member
- abstract - instance
- whole - part
- process - step
- object - attribute

- generalization - specific

The effect : R (the reader) recognizes the situation presented in S as providing additional details for N. R identifies the element of subject matter for which detail is provided.

Locus of the effect : N and S

(Mann et Thompson, 1987, p. 52)

La définition de la relation d'*Élaboration* dans la RST met en avant le pôle de la différence. Des moyens sont donnés à l'analyste pour définir la relation d'*Élaboration* entre deux segments en termes de six couples *segment élaboré/segment élaborant*. Par exemple, pour le couple *process/step*, le segment élaboré décrit un *procès* et le segment élaborant décrit une *étape de ce procès*.

Mann et Thompson (1986, p. 64) donnent des exemples pour certains types de la relation déjà présents dans cette version de la RST :

- **Set - Member** (Ensemble - Membre)

- (21) I love to collect classic automobiles. My favorite car is my 1899 Duryea.
J'adore collectionner les voitures classiques. Ma voiture préférée est ma Duryea 1899.

- **Whole - Part** (Tout - Partie)

- (22) Karen is so photogenic. Her smile is perfect.
Karen est trop photogénique. Son sourire est parfait.

- **Process - Step** (Procès - Étape)

- (23) It's time to make our cake. I'm going to take out the milk and eggs.
Il est l'heure de faire le gâteau. Je vais sortir le lait et les oeufs.

- **Object - Attribute** (Objet - Attribut)

- (24) I'm Officer Jordan. I was born in 1952 and I joined the police force in 1970.
Je suis l'officier Jordan. Je suis né en 1952 et j'ai rejoint les forces de police en 1970.

- **Generalization - Specific** (Généralisation - Spécifique)

- (25) Your performance distresses me. You come in drunk and you insult the busboy.
Ton comportement me désole. Tu rentres saoul et tu insultes l'aide-serveur.

Mann et Thompson (1987) illustrent la relation d'*Élaboration* avec l'exemple suivant :

- (26) (a) Sänga-Säby-Kursgård, Sweden, will be the site of the 1969 International Conference on Computational Linguistics.
 (b) It is expected that some 250 linguists will attend from Asia, West Europe, East Europe including Russia, and the United States.
 (c) The conference will be concerned with the application of mathematical and

computer techniques to the study of natural languages, the development of computer programs as tools for linguistic research, and the application of linguistics to the development of man-machine communication systems.

(a) *Sånga-Säby-Kursgård, Suède, sera le lieu où se tiendra la Conférence Internationale de 1969 en linguistique computationnelle* (b) *250 linguistes sont attendus d'Asie, d'Europe de l'Ouest, d'Europe de l'Est incluant la Russie et les Etats-Unis.* (c) *La conférence portera sur les applications en mathématiques et techniques computationnelles pour l'étude du langage naturel, le développement de systèmes de communication homme/machine.* (Mann et Thompson, 1987, p. 53)

Selon Mann et Thompson (1987), le but de ce petit discours est d'avertir les lecteurs de la tenue d'une conférence. Le premier segment donnant l'information essentielle dans ce discours, i.e. la tenue de la conférence «the International Conference on Computational Linguistics», est considéré comme le noyau (représenté dans la figure 2.2 par une barre verticale). Les segments (b) et (c) sont tous deux des satellites reliés au noyau (a) par la relation d'*Élaboration*. Ils apportent des détails supplémentaires sur l'entité «Conference». Les auteurs spécifient qu'il s'agit ici d'une *Élaboration* du type *object-attribut* (*objet-attribut*). L'objet est la «conférence» qui constitue l'élément essentiel introduit par le noyau et les attributs sont le nombre de linguistes attendu à la conférence (b) et les disciplines présentes à la conférence (c).

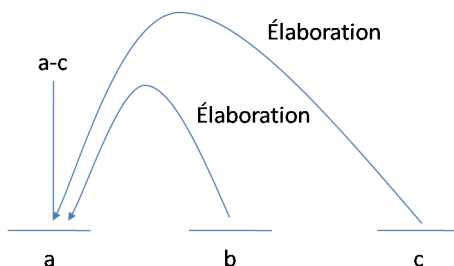


FIG. 2.2 – Structure de (26) en RST

Un certain nombre de problèmes demeurent avec cette définition.

Premièrement, cette définition semble indiquer qu'il n'existe pas une *Élaboration* mais plusieurs relations (une pour chaque sous-type). Il est légitime de se demander si nous n'avons pas six relations distinctes, nommées et identifiées, au lieu d'une. Les auteurs de la RST ne distinguent pas six relations car ces derniers signalent qu'un ou des sous-types de la relation peuvent être mis en jeu («*in one or more of the ways listed below*»). Nous verrons, dans la section suivante, la solution proposée par Hovy et Maier (1991) qui consiste à établir une taxonomie de relations, allant d'une relation générale à des sous-types de plus en plus précis.

Les auteurs de la RST font également remarquer qu'il n'est pas toujours facile de distinguer ces sous-types : «*the subtypes are not treated as separated relations because in many cases in text they are not distinguishable*» (Mann, William C. and Taboada, Maite, 2007). Cette remarque met au jour un second problème. Il nous semble effectivement qu'il est

difficile de distinguer ces sous-types, d'autant qu'aucune définition des sous-types n'est disponible. En appliquant la définition de l'*Élaboration* à notre exemple (11), nous pouvons relier les deux phrases par la relation d'*Élaboration* mais nous ne savons pas avec exactitude quel est le ou les sous-types mis en oeuvre parmi *whole/part*, *process/step* et *generalization/specific*.

(11) (a) Paul s'est promené dans le parc. (b) Il a traversé le petit bois jusqu'à la rivière.

Comme dans les définitions de la Grammaire Fonctionnelle, la RST tient compte du fait que l'*Élaboration* peut relier soit deux segments dans leur globalité, soit un segment et un élément du segment précédent. Il nous semble que certains sous-types de la relation vont alors s'appliquer soit à la *situation* comme dans (11) où celle-ci décrite dans le segment (a) fait l'objet de l'élaboration tandis que d'autres sous-types vont s'appliquer à une *entité* du segment tels que *object/attribute* dans (26) où l'élaboration porte sur une entité, la «Conférence» introduite dans le segment (a). Cependant, la définition n'apporte pas d'éléments clairs sur ce sous-type.

Contrairement aux définitions de la Grammaire Fonctionnelle qui autorisent l'élaboration d'un grand nombre d'élément du premier segment, la RST ne va pas relier les segments des exemples (9) et (10) avec la relation d'*Élaboration* car le lecteur va reconnaître dans le deuxième segment de (9) l'intention de justifier le contenu du premier segment et dans le deuxième segment de (10) l'intention de donner un cadre au contenu du premier segment. La RST propose la relation de *Justification* pour (9) et la relation de *Circonstance* pour (10).

(9) (a) Paul s'est promené dans le parc. (b) Il n'arrivait plus à travailler.

(10) (a) Paul s'est promené dans le parc. (b) Le soleil commençait à se coucher au bord du lac.

Enfin, le sous-type d'*Élaboration process - step* serait proposé par la RST pour l'exemple (12) :

(12) (a) Mixel a escaladé le Vignemale hier matin. (b) Il est parti du refuge des Oulettes au lever du jour, (c) puis il a passé la Hourquette d'Ossoue vers 9h, (d) et il est arrivé au sommet vers midi.

Pour conclure, il nous semble que la définition de la relation d'*Élaboration* et ses six sous-types correspondent à l'intuition que nous avons de la notion d'élaboration. D'abord nous sommes d'accord sur le fait qu'une élaboration peut s'appliquer à une situation ou à une entité introduite dans un segment. Puis nous sommes également d'accord sur le fait que la notion d'élaboration se divise en plusieurs sous-types (exemplification, ajout de détails sur un objet, description d'une étape...). Toutefois, il nous semble que des définitions plus détaillées et des indications sur la mise en oeuvre de chaque sous-type augmenteraient l'applicabilité de la définition.

Le pôle de l'identité, tel que décrit par Kleiber et Vassiliadiou (2009), concerne deux autres relations de discours, la *Reformulation (Restatement)* et le *Résumé (Summary)*. La

Reformulation et le *Résumé* sont toutes les deux des relations de reformulation mais avec un recouvrement différent : *Reformulation* relie deux segments qui recouvrent globalement le même message tandis que *Résumé* relie deux segments lorsque le satellite recouvre une plus petite partie du message que le noyau. Nous discuterons ici seulement la définition de la relation de *Reformulation* :

Constraints on the Nucleus : None

Constraints on the Satellite : None

Constraints on the N+S combination The Satellite restates the Nucleus, where S (Satellite) and N (Nucleus) are comparable bulk.

The effect : R (the reader) recognizes S as a restatement of N.

Locus of the effect : N and S

(Mann et Thompson, 1987, p. 71)

Mann et Thompson (1987) accompagnent cette définition de l'exemple suivant extrait d'une publicité :

- (27) (a) A WELL-GROOMED CAR REFLECTS ITS OWNER. (b) The car you drive says a lot about you. (Mann et Thompson, 1987, p. 71)
 (a) Une voiture bien entretenue reflète son propriétaire. (b) La voiture que vous conduisez en dit beaucoup sur vous.

Effectivement, nous reconnaissons (a) et (b) comme ayant un contenu comparable. Mais (b) est également compris comme une généralisation de (a). La prédication est la même : «refléter son propriétaire» et «en dire beaucoup sur son propriétaire» ; mais le sujet de la prédication est différent : dans (a) nous avons «une voiture entretenue» tandis que dans (b) nous avons un sujet plus général qui englobe celui de (a) «la voiture que vous conduisez». Est-ce que ce point n'est pas suffisant pour considérer que ces deux segments sont reliés par la relation d'*Élaboration general - specific* ?

Des questions se posent sur cette relation de *Reformulation*. L'existence de reformulations pures dans les discours nous semble improbable. Un locuteur ou un scripteur peut éventuellement reformuler un message, mais en adoptant un point de vue différent, en clarifiant son message, en ajoutant des détails... Tous ces cas peuvent être traités avec la relation d'*Élaboration*. D'ailleurs, la reformulation est souvent considérée comme un sous-type d'*Élaboration* (Hovy et Maier, 1991).

La définition de la relation d'*Élaboration* dans la RST rend compte des diverses façons d'élaborer dans le discours. Les travaux que nous allons présenter dans la section suivante prennent directement appui sur ces travaux, auxquels ils ajoutent deux nouveautés : ils organisent les six sous-types de la RST dans une hiérarchie de spécificité croissante et incluent la reformulation comme un sous-type d'*Élaboration*.

2.2.2.3 La relation d'*Élaboration* chez Hovy et Maier (1991)

La synthèse de Hovy et Maier (1991) est similaire à celle que nous proposons dans ce chapitre dans le sens où les auteurs ont considéré les différentes définitions qui ont été

données à la relation d'*Élaboration* ou à des notions proches de celles-ci dans plusieurs théories, cf. section 2.1.

Les objectifs que nous poursuivons dans un premier temps dans ce chapitre diffèrent toutefois de ceux de Hovy et Maier (1991) : nous souhaitons faire un inventaire de toutes les façons de définir la relation d'*Élaboration* en essayant d'établir les correspondances et/ou les divergences d'une théorie à une autre, tandis que l'objectif de Hovy et Maier (1991) est d'aboutir à leur propre définition de la relation d'*Élaboration* en comparant les définitions et en cherchant à les fusionner lorsque celles-ci sont semblables. Nous proposerons notre propre définition à la fin du chapitre suivant.

Leur définition s'appuie essentiellement sur celles données à la relation par la Grammaire Fonctionnelle et la RST :

One text segment expands on the other by specifying it in greater detail or specifying it in other words, according to one of the following ways :

- set-member (ensemble-membre)
- process-step (procès-étape)
- part-whole (partie-tout)
- object-attribute (objet-attribut)
- abstract-instance (abstrait-exemple)
- general-specific (général-spécifique)
- restatement (reformulation)

(Hovy et Maier, 1991, p. 8)

Cette définition présente des similitudes importantes avec celle de la RST et celle de Halliday (1985). L'*Élaboration* est une relation d'expansion au sens de Halliday (1985) : un segment spécifie le contenu d'un segment précédant en donnant des détails selon une ou plusieurs «façons» inspirées de la définition de la RST. Hovy et Maier (1991) ont inclus la relation de *Reformulation*, qui est traitée comme une relation à part entière dans la RST, se rapprochant ainsi de la conception de Halliday (1985), et la relation d'*Identification*¹¹.

Hovy et Maier (1991) ont analysé approximativement quatre cents relations parmi les travaux d'une trentaine de chercheurs pour finalement aboutir à une taxonomie de 70 relations organisées dans une hiérarchie de spécificité croissante. Plus une relation se trouve haut dans la hiérarchie, plus celle-ci a une acception générale. Plus une relation se trouve bas dans la hiérarchie, plus celle-ci est spécifique (avec des traits de définition supplémentaires). La figure 2.3 donne un aperçu de cette taxonomie concernant la relation d'*Élaboration*. La relation d'*Élaboration* domine cinq sous-relations intermédiaires : *Restatement*, *Identification* et les trois relations *ElabObject*, *ElabPart* et *ElabGenerality* qui dominent à leur tour les relations décrites dans la RST : *ElabObject* domine *Object Attribute*, *Object-Function* et *Set-Member* pour lesquels l'élaboration porte sur un objet ; *ElabPart* domine *Whole-Part* et *Process-Step* pour lesquels l'éventualité élaborante constitue une partie de l'éventualité élaborée. D'après les exemples de Mann et Thompson (1986), *Whole-Part* porte sur des états (22) tandis que *Process-Step* porte sur des événements (23) :

(22) Karen is so photogenic. Her smile is perfect.

¹¹ *Identification* est une relation permettant d'établir une référence et de la maintenir (Grimes, 1975). Elle ne constitue pas de notre point de vue une relation de discours.

(23) It's time to make our cake. I'm going to take out the milk and eggs.

Enfin *ElabGenerality* domine *General-Specific* et *Abstract-Instance*.

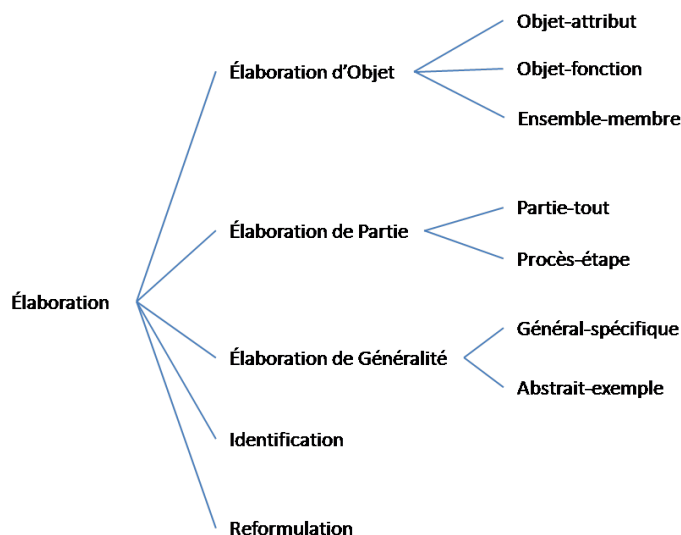


FIG. 2.3 – Extrait de la taxonomie de Hovy et Maier (1991)

Le reproche qui a été fait aux approches dites multiplicatrices (cf. chapitre 1, section 1.2), telles que la RST, peut être donné également à cette solution. Autrement dit, les noeuds terminaux (les feuilles) de la taxonomie peuvent à nouveau être subdivisés, les relations pouvant être toujours déterminées avec plus de précision. Mais cela ne représente pas un problème du point de vue de Hovy et Maier (1991) car leur taxonomie propose des noeuds intermédiaires qui fournissent des relations de discours fonctionnelles pour donner une représentation de la structure discursive d'un texte et les niveaux inférieurs, avec leur précision sémantique, peuvent trouver leur utilité selon une application donnée. De plus, le niveau intermédiaire qu'apportent Hovy et Maier (1991) permet d'affiner la définition de la relation d'*Élaboration* de la RST en distinguant les élaborations qui portent sur des objets, les élaborations qui portent sur des événements et des élaborations plus globales allant du général au particulier.

2.2.2.4 Bilan

Les approches que nous avons présentées ci-dessus laissent une place importante à l'intuition et à l'interprétation des lecteurs ou allocutaires. Halliday (1985) suppose que la reconnaissance par un lecteur ou un allocutaire de la relation d'*Élaboration* entre deux propositions est un indice suffisant pour considérer l'ensemble de deux propositions juxtaposées comme une proposition complexe. Dans la RST, les relations de discours sont reconnues par le lecteur ou l'allocutaire selon les intentions du scripteur ou du locuteur et les effets que celui-ci souhaite produire.

Les approches que nous souhaitons maintenant présenter dans les sections suivantes ont cherché à donner des règles formelles permettant d'inférer les relations de discours, parmi lesquelles la relation d'*Élaboration*.

2.2.3 Approches formelles

Le passage aux approches formelles est une étape importante dans la définition des relations de discours, et en particulier de la relation d'*Élaboration*. La formalisation répond à la nécessité d'expliquer le fonctionnement du discours au moyen d'un système de règles. Nous proposons de présenter dans cette section les définitions de la relation dans deux théories formelles - la théorie de Hobbs (1990) et la théorie de Kehler (2002) - toutes deux issues de l'Intelligence Artificielle. Leur objectif est de donner les moyens à un système informatique de reproduire les inférences produites par les êtres humains lors de l'interprétation d'un discours. Enfin, dans le chapitre suivant, nous focaliserons notre attention sur une autre théorie formelle, la SDRT (Asher, 1993; Asher et Lascarides, 2003) que nous avons choisie comme cadre d'analyse.

2.2.3.1 La relation d'*Élaboration* chez Hobbs (1979, 1985, 1990)

Hobbs (1979, 1985, 1990) recourt aux inférences pour définir les relations de cohérence.

La relation d'*Élaboration* fait partie de la famille des relations nommée *Expansion*. Hobbs stipule que le lecteur doit établir la même inférence à partir des deux propositions reliées S_0 et S_1 pour pouvoir inférer *Élaboration*.

Elaboration :

S_1 is an Elaboration of S_0 if a proposition P follows-from the assertions of both S_0 et S_1 (but S_1 contains a property of one of the elements of P that is not in S_0) (Hobbs, 1979, p. 73)

Elaboration (S_0, S_1) :

Infer the same proposition P from the assertions of S_0 and S_1 (Hobbs, 1990, p. 95)

La définition de Hobbs (1990) présente une évolution : la suppression de la parenthèse (qui spécifie à l'origine une exigence de nouveauté) entraîne un élargissement de la définition qui couvre alors plus de cas. Les deux définitions apportent la même information : généralement, la proposition élaborante apporte des informations nouvelles mais la seconde définition permet également d'inclure les cas de répétition et de reformulation : «Frequently, the second segment adds crucial information, but this is not specified in the definition since it is desirable to include pure repetitions under the headline of Elaboration», (Hobbs, 1990, p. 96).

Chez Hobbs (1990), les deux segments reliés par la relation d'*Élaboration* ont des contenus similaires avec, pour les cas typiques, un ajout d'informations, comme dans l'exemple suivant :

- (28) (a) Go down First Street. (b) Just follow First Street down three blocks to Adams Street. (Hobbs, 1990, p. 96)
 (a) *Continuez à descendre sur First Street.* (b) *Suivez First Street en descendant encore trois blocks jusqu'à Adams Street.*

Pour analyser correctement ce discours, Hobbs fait remarquer qu'il est nécessaire d'inférer la même proposition P à partir des deux assertions (a) et (b) (et ne pas interpréter (b) comme une nouvelle instruction). La proposition P commune à (a) et (b) est une assertion du type «se rendre à un certain point de First Street». Pour construire cette inférence commune, il faut comprendre que l'éventualité «follow» (suivre) implique «go down» (se rendre) et faire correspondre les deux trajets évoqués dans les deux assertions : «First Street». Enfin, nous notons que la deuxième assertion apporte des informations supplémentaires sur la première assertion en précisant le point final du trajet «Adams Street» et la distance de ce trajet «three blocks». Cette analyse est reproduite ci-dessous :

- (a) go(Agent : you ; But : X ; Trajet : First Street ; Distance : Y)
 (b) go(Agent : you ; But : Adams Street ; Trajet : First Street ; Distance :
 Three Blocks)
 (Hobbs, 1990, p. 96)

(b) restreint l'interprétation de (a), c'est-à-dire que l'interprétation globalisante possible de (a) «descendre toute la First Street» est totalement exclue à la lecture de (b). Cette analyse renforce l'idée de Hobbs sur la relation d'*Élaboration*, à savoir que celle-ci a pour fonction principale d'éviter les malentendus ou le manque d'information.

Les définitions logiques de Hobbs (1979, 1990) (*Elaboration*(a, b) si $a \rightarrow P \wedge b \rightarrow P$) ne sont pas immédiatement transparentes par rapport à l'intuition que l'on peut avoir de l'*Élaboration*. Ces définitions sont destinées à montrer que la proposition élaborante ne sort pas de la situation dénotée par la proposition élaborée. Cette information est représentée par P . Des détails sur P peuvent être apportés par la proposition élaborante. Cette définition est tournée vers l'identité¹². Le critère mis en avant dans la définition est ce qui unit les propositions reliées par la relation et non ce qui les sépare : «*At a sufficient deep level, the two sentences say the same thing*» (Hobbs, 1979, p. 73).

Hobbs (1990) cherche à inclure, dans une seule et même définition, des cas d'*Élaboration* avec apport d'informations supplémentaires et des cas d'*Élaboration* semblables à de la répétition ou de la reformulation. Kleiber et Vassiliadiou (2009) font remarquer que ces deux cas présentent des différences structurelles. Les premiers cas peuvent être décrits en termes d'implication unilatérale tandis que les seconds peuvent être décrits en termes d'équivalence logique : si on a *Elaboration*(a, b) dans le premier cas, nous avons $b \rightarrow a$ et dans le deuxième cas, nous avons $a \leftrightarrow b$.

Kleiber et Vassiliadiou (2009) rejettent la définition de Hobbs (1990) (*Elaboration*(a, b) si $(a \rightarrow P \wedge b \rightarrow P)$) et proposent que les exemples d'*Élaboration* tels que (28) soient caractérisés de façon plus précise en établissant une relation d'implication unilatérale entre les deux propositions (a) et (b), c'est-à-dire que (b) implique (a) mais pas l'inverse ($b \rightarrow a$ et

¹²Hobbs considère la relation d'*Élaboration* comme un cas particulier de la relation *Parallèle*.

$a \nrightarrow b$). Dans l'exemple (28), «descendre sur First Street durant trois blocs jusqu'à Adams Street» implique «descendre First Street» mais «descendre First Street» n'implique pas «descendre sur First Street durant trois blocs jusqu'à Adams Street» car effectivement on peut s'arrêter avant d'arriver à Adams Street tout en pouvant inférer qu'on a descendu First Street (avec une interprétation atélisque de «go down» (descendre)).

Hobbs ne retient évidemment pas l'implication unilatérale $b \rightarrow a$ pour déclencher la relation d'*Élaboration* car cette formulation ne permet pas d'inclure les cas de reformulation et de répétition. Dans ces cas particuliers, S0 implique S1 et S1 implique S0, autrement dit S0 et S1 sont dans une relation d'équivalence logique.

La proposition de Kleiber et Vassiliadiou (2009) (*Elaboration*(a, b) si $b \rightarrow a$) entraîne la subordination de S1 à S0, tandis que la proposition de Hobbs (1990) place les deux propositions (a) et (b) reliées par la relation d'*Élaboration* à un même niveau.

Si on peut analyser l'exemple (28) en termes d'implication unilatérale, toutes les élaborations ne peuvent pas être traitées ainsi. Les définitions inférentielles présentées ci-dessus ne permettent pas de rendre compte des élaborations du type *process - step* décrites dans la RST :

- (29) (a) Mixel a escaladé le Vignemale. (b) Il est parti du refuge des Oulettes au lever du jour.

Dans l'exemple ci-dessus, la phrase (b) décrit un événement qui est une étape de l'événement décrit dans la phrase (a). En nous appuyant sur la définition de Hobbs (1990), nous ne pouvons pas inférer *Élaboration* étant donné que nous ne pouvons pas inférer de proposition commune P de (a) et de (b). La caractérisation plus précise de Kleiber et Vassiliadiou (2009) ne tient pas non plus étant donné que dans ce discours, nous n'avons pas $b \rightarrow a$. Pourtant, nous considérons bien cet exemple comme un cas d'*Élaboration*.

La définition de la relation d'*Élaboration* chez Hobbs (1979, 1990) couvre des cas de redescription d'une même éventualité. Hobbs peut donc rendre compte, avec sa définition d'*Élaboration*, d'un exemple comme (12) à condition que toutes les étapes de l'événement décrit dans la proposition élaborée soient données dans la ou les propositions élaborantes :

- (12) (a) Mixel a escaladé le Vignemale hier matin. (b) Il est parti du refuge des Oulettes au lever du jour, (c) puis il a passé la Hourquette d'Ossoue vers 9h, (d) et il est arrivé au sommet vers midi. (Bras, 2007, p. 232)

L'ensemble des trois étapes réunies (b), (c) et (d) constituent une redescription de l'événement «escalader le Vignemale». Il est donc possible, dans le cadre de la théorie de Hobbs, d'inférer *Élaboration* entre (a) et la réunion de (b), (c) et (d) liés par la relation *Occasion*¹³ (Hobbs, 1990). La figure 2.4 représente la structure de (12) que nous construisons au moyen des relations de Hobbs (1990).

¹³Deux configurations rendent possible l'inférence de la relation *Occasion* chez Hobbs (1990) : 1. Un changement d'état peut être inféré de la proposition S0, dont l'état final peut-être inféré de la proposition S1. 2. Un changement d'état peut être inféré de la proposition S1, dont l'état initial peut être inféré de S0.

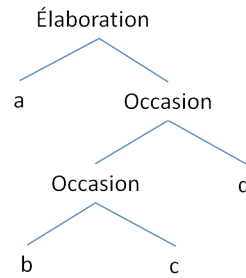


FIG. 2.4 – Structure de (5) chez Hobbs

La définition de Hobbs (1990) permet donc également de rendre compte de l'exemple (11) si le segment (b) est considéré comme une redescription de l'événement décrit dans le segment (a). De (a) et (b), on peut inférer la proposition P «Paul s'est déplacé dans le parc».

(11) (a) Paul s'est promené dans le parc. (b) Il a traversé le petit bois jusqu'à la rivière.

En revanche, il n'est pas possible d'inférer de propositions communes à partir de (a) et (b) dans les exemples (9) et (10), ce qui permet de prédire que les propositions (a) et (b) ne sont pas reliées par la relation d'*Élaboration*.

(9) (a) Paul s'est promené dans le parc. (b) Il n'arrivait plus à travailler.

(10) (a) Paul s'est promené dans le parc. (b) Le soleil commençait à se coucher au bord du lac.

Nous souhaitons maintenant discuter un autre exemple de Hobbs analysé avec la relation d'*Élaboration* :

(30) (a) John can open Bill's safe. (b) He knows the combination. (Hobbs, 1990, p. 96)
 (a) John peut ouvrir le coffre-fort de Bill. (b) Il connaît la combinaison

L'analyse de Hobbs est la suivante : de (a) et (b), on infère P «Bill connaît une action qui lui permet d'ouvrir le coffre-fort de Bill». Des connaissances du monde nous permettent d'établir le lien entre la modalité exprimée par «can» (pouvoir) et le fait d'avoir une connaissance particulière. L'*élaboration* porte sur cette connaissance qui est la mémorisation de la combinaison du coffre-fort.

Cette analyse est discutable car d'autres analystes proposent ici une *Explication*, c'est-à-dire que la connaissance de la combinaison rend John possiblement capable d'ouvrir le coffre-fort. Cette suggestion est de plus appuyée par le fait que le marqueur *because*, prototypique de l'*Explication*, peut être inséré entre les deux propositions.

Selon Hobbs, dans (30), l'*Élaboration* et l'*Explication* se fondent en quelque sorte l'une dans l'autre. La reconnaissance d'*Explication* se base sur l'inférence de (b) comme étant une cause possible de (a). La reconnaissance d'*Élaboration* se base sur la double inférence de P à partir des assertions de (a) et (b). Très souvent, la proposition P est l'assertion de

(a). Dans l'exemple (30), la proposition P «Bill connaît une action qui lui permet d'ouvrir le coffre-fort de Bill» n'est autre que (a), et donc (b) implique (a) dans ce cas. Selon Hobbs (1990), il est probable que cette implication soit perçue comme une causalité :

- (a) \rightarrow P
- (b) \rightarrow P
- (a) = P
- (b) \rightarrow a

(30) est un exemple de la confusion entre les relations de discours *Élaboration* et *Explication* sur laquelle nous reviendrons dans cette thèse dans le chapitre 9.

Selon Hobbs, l'exemple (30) est un cas d'*Élaboration* appelée *function-structure* (fonction-structure). Le premier segment décrit une éventualité comme étant une fonction (dans (30) la capacité d'ouvrir le coffre-fort) et le deuxième segment détaille la structure interne de cette éventualité (dans (30) la connaissance de la combinaison du coffre-fort). L'exemple suivant est également un exemple de ce type d'*Élaboration* extrait d'une description d'un algorithme :

- (31) (a) Initialize T. (b) Set stack A empty and set link variable P to T. (Hobbs, 1990, p. 97)
 (a) Initialise T. Vide la pile A et lie la variable P à T.

Selon l'analyse de Hobbs, le segment (a) décrit la fonction (ou le rôle) que jouent les instructions dans le programme. Le segment (b) décrit spécifiquement les deux instructions à suivre : «Set stack empty» et «link variable P to T», deux propositions par ailleurs reliées par la relation *Parallèle* selon les définitions de Hobbs.

Nous proposons, pour terminer cet exposé, l'analyse de trois autres exemples de Hobbs (1990) :

- (32) (a) This immense tract of time is only sparsely illuminated by human relics. (b) Not enough material has yet been found for us to trace the technical evolution of East Asia. (Hobbs, 1990, p. 97)
 (a) Cette immense période de temps n'est que partiellement mise en lumière par des reliques humaines. (b) Nous n'avons pas encore trouvé assez de matériel pour nous permettre de retracer l'évolution technique en l'Asie de l'Est.

De la première phrase, on peut inférer la proposition P suivante «nous manquons d'éléments pour étudier cette période de temps» à partir de «sparsely illuminated» (partiellement mise en lumière). De la deuxième phrase, on peut inférer la même proposition P à partir de «not enough material» (pas assez de matériel). Afin d'obtenir cette même proposition P à partir des deux segments : «il manque du matériel pour étudier cette période de temps», il faut faire correspondre «relics» (reliques) du segment (a) et «material» (matériel) du segment (b) comme renvoyant aux mêmes objets et faire également correspondre «this immense tract of time» (longue période) dans (a) et «the technical evolution of East Asia» (l'évolution technique de l'est de l'Asie) dans (b) comme renvoyant à la même période de temps.

- (33) (a) Generally blood donor quality is held high by avoiding commercial donors...
 (b) Extremely careful selection of paid donors may provide safe blood sources in some extraordinary instances, but generally it is much safer to avoid commercially obtained blood. (Hobbs, 1990, p. 97)
(a) Généralement la qualité des dons de sang est plus élevée si on évite les donneurs commerciaux... (b) Une sélection extrêmement rigoureuse des dons financés pourrait fournir des sources sûres de sang, dans quelques cas exceptionnels, mais généralement il est plus sûr d'éviter le sang obtenu de façon commerciale.

Selon Hobbs, on peut inférer à partir du premier segment (a) la proposition P suivante «une qualité élevée des donneurs de sang est une façon de minimiser les risques»¹⁴ à partir de «blood donor quality is held high» (la qualité des donneurs de sang est plus élevée). Dans la deuxième phrase, on peut inférer la même proposition P à partir de «safe blood sources» (source de sang sûre) et «much safer» (plus sûr) qui impliquent la minimisation des risques.

- (34) (a) Time is running out on Operation Condor. (b) Nixon must clear himself by early in the new year or lose his slipping hold on the party. (Hobbs, 1990, p. 98)
(a) Le temps s'écoule pour l'Opération Condor. (b) Nixon doit s'expliquer avant le début de l'année sous peine de perdre sa place au soleil.

Dans cet exemple, la reconnaissance de la relation d'Élaboration est basée sur la possibilité d'inférer que «time is running out» (le temps manque) et «must... be early in the new year» (doit... avoir lieu en ce début de nouvelle année) renvoient à la même idée, c'est-à-dire «qu'on attend rapidement une réaction de M. Nixon». Cette inférence se base également sur le fait que le lecteur sait que *Nixon* était impliqué dans l'Opération Condor.

Hobbs (1990) distingue, en plus de la relation d'Élaboration, la relation d'Exemplification.

Exemplification (S0,S1) :

Infer P(A) from the assertion of (S0) and P(a) from the assertion of (S1), where a is a member or subset of A. (Hobbs, 1990, p. 98).

Nous allons considérer l'Exemplification en examinant les deux exemples suivants et leur analyse proposée par Hobbs (1990).

- (35) (a) This algorithm reverses a list. (b) If its input is «(A, B, C)», its output is «(C, B, A)». (Hobbs, 1990, p. 98)
(a) Cet algorithme inverse une liste. (b) Si son entrée est «A, B, C», sa sortie est «C, B, A».

Pour inférer la relation d'Exemplification, il faut d'abord inférer de «reverses» dans (a) une relation causale entre l'entrée et la sortie de l'algorithme : «le résultat Y (output) de l'algorithme est l'inverse de X (input)», puis reconnaître que «A, B, C» est une liste d'entrée (input) de l'algorithme et «C, B, A» est la sortie (output) de l'algorithme.

¹⁴Il nous semble plus exact d'inférer «la qualité des donneurs de sang est plus élevée si on évite les donneurs commerciaux» pour (a) et (b). Mais cela ne change pas considérablement l'analyse.

L'exemple suivant nous amène à discuter la prise en compte de cette relation d'*Exemplification* en plus de la relation d'*Élaboration* :

- (36) (a) We cannot affirm that the technical evolution of East Asia followed the same course as it did in the West. (b) Certainly no stage corresponding to the Mousterian tradition has been found in China. (Hobbs, 1990, p. 99)
 (a) *Nous ne pouvons pas affirmer que l'évolution technique de l'Asie de l'Est a suivi le même cours que dans l'Ouest.* (b) *Assurément aucune étape correspondant à la tradition Mousteriane n'a été trouvée en Chine.*

L'analyse de Hobbs est la suivante : on établit une correspondance entre «cannot affirm» (ne peut pas affirmer) et «no stage... has been found» (aucune étape... n'a été trouvée). L'inférence s'appuie également sur des connaissances du monde, à savoir que la *Chine* fait partie de l'*Asie de l'Est* et que la *tradition Mousteriane* est une période de l'*évolution technique de l'Asie de l'Ouest*.

La limite entre *Élaboration* et *Exemplification* ne semble pas claire au vu de l'analyse de l'exemple (36). Il nous semble tout à fait justifiable d'inférer *Élaboration* entre (a) et (b) dans (36) avec la proposition P commune : «Nous ne pouvons pas établir de correspondance entre l'évolution technologique de l'Asie de l'Est et celle de l'Asie de l'Ouest». Pour inférer *Élaboration*, nous nous basons sur les mêmes inférences décrites ci-dessus. Nous considérons donc la relation d'*Exemplification* comme un cas particulier de l'*Élaboration* lorsque les inférences s'appuient sur des relations d'inclusion entre les éléments contenus dans S0 et S1 : «la *Chine* est une part de l'*Asie de l'Est*» ; «la *tradition Mousteriane* est une période de l'*évolution technique de l'Asie de l'Ouest*».

Conclusion Les relations d'*Élaboration* et d'*Exemplification* chez Hobbs ne permettent de rendre compte que des cas d'*Élaboration* qui impliquent une redescription d'une même éventualité (*Élaboration*) ou le passage d'une généralisation à une spécification (*Exemplification*). Tous les autres types d'élaboration qui sont décrits dans les autres théories ne peuvent pas être analysés dans la théorie de Hobbs, comme par exemple les élaborations de type *process-step* dont nous avons donné un exemple dans cette section, mais également les élaborations du type *object-attribute*, *object-function* et *whole-part*. Les travaux que nous allons présenter dans la prochaine section sont directement inspirés des travaux de Hobbs (1990).

2.2.3.2 La relation d'*Élaboration* chez Kehler (2002, 2004)

Kehler (2002, 2004) propose une théorie des relations de discours basée sur les travaux de Hobbs (1979, 1990). Il regroupe les mêmes relations (*Parallèle*, *Élaboration*, *Exemplification*, *Généralisation*...) dans la famille de relations nommées *RESSEMBLANCE* et propose des définitions très similaires pour chacune d'entre elles. Toutefois il ajoute dans les définitions la prise en compte d'un ensemble d'entités. La relation d'*Élaboration* est un cas particulier de la relation de *Parallèle* dans laquelle les deux segments reliés par la relation décrivent en fait la même éventualité.

(Elaboration) Infer $p(a_1, a_2)$ from the assertion of S_1 et S_2 . (Kehler, 2002, p. 18)

Kehler (2002, 2004) met également en avant dans sa définition l'identité des deux segments reliés par la relation : «Elaborations are generally restatments» (Kehler, 2002, p. 18), en insistant sur le fait que l'on infère une même proposition P à partir de S_1 et S_2 et qu'un ensemble d'entités (i.e. référents) obtenu à partir des deux propositions S_1 et S_2 sont identiques. Selon Kehler, la différence entre les deux segments provient du point de perspective choisi ou du niveau de détail auquel les entités sont décrites.

- (37) (a) A young aspiring politician was arrested in Texas today. (b) John Smith, 34, was nabbed in a Houston law firm (c) while attempting to embezzle funds for his campaign. (Kehler, 2002, p. 18)
 (a) *Un jeune aspirant à devenir politicien a été arrêté au Texas aujourd'hui. (b) John Smith, 34 ans, s'est fait pincer dans un cabinet d'avocat à Houston (c) alors qu'il tentait de détourner des fonds pour sa campagne.*

Lorsqu'un lecteur interprète (37), celui-ci comprend qu'un seul et même événement est décrit dans les deux propositions, i.e. «l'arrestation de John Smith». Il comprend que les expressions référentielles «A young aspiring politician» dans la première phrase et «John Smith, 34» dans la seconde renvoient au même référent et que les entités correspondant aux prédicats des deux phrases, respectivement «was arrested» et «was nabbed» renvoient à la même action. Les informations nouvelles sont apportées avec la localisation plus précise de l'événement, s'appuyant sur des connaissances du monde, à savoir que «Houston» se situe dans le «Texas» et par une explicitation des circonstances de l'arrestation «while attempting to embezzle funds for his campaign», que l'on peut également comprendre comme étant les raisons de l'arrestation. Toutes ces inférences permettent d'établir que la relation d'*Élaboration* relie les segments (a) et (b-c).

- (38) (a) The new Republican president took a swipe at abortion in his first week of office. (b) In a White House ceremony yesterday, George W. Bush signed an executive order banning support to international agencies which offer abortion as one of their services. (Kehler, 2004, p. 250)
 (a) *Le nouveau président républicain a porté un rude coup à l'avortement dès la première semaine de son mandat. (b) Hier lors d'une cérémonie à la Maison Blanche, George W. Bush a signé un ordre exécutif supprimant les aides financières aux agences internationales qui proposent la possibilité d'avorter.*

De même pour (38), lorsqu'un lecteur interprète ce discours, celui-ci comprend qu'une seule et même action est décrite dans les deux propositions, i.e. «Le président a porté un rude coup à l'avortement». Il comprend que les expressions référentielles «The new Republican president» du segment (a) et «George W. Bush» du segment (b) renvoient au même référent. Les informations nouvelles apportées par le segment (b) concernent à la fois la localisation «In a White House ceremony» et l'explicitation de «took a swipe at abortion» par «signed an executive order banning support to international agencies which offer abortion as one of their services».

Kehler (2002, 2004) conserve également la relation d'*Exemplification* de Hobbs (1990).

Exemplification : Infer $p(a_1, a_2, \dots)$ from the assertion of S_1 and $p(b_1, b_2, \dots)$ from the assertion of S_2 , where b_i is a member or subset of a_i for some i . (Kehler, 2002, p. 17)

Nous allons voir en quoi l'exemple de la relation d'*Exemplification* (39) diffère de l'exemple (38) d'*Élaboration* :

- (39) (a) Republican presidents often seek to put limits on federal funding of abortion.
 (b) In his first week of office, George W. Bush signed a ban on contributing money to international agencies which offer abortion as one of their services. (Kehler, 2004, p. 249) (a) *Les présidents républicains cherchent souvent à imposer des limites aux fonds donnés par l'état fédéral en faveur de l'avortement. Lors de la première semaine de son mandat, George W. Bush a signé un décret interdisant toute contribution financière en faveur des agences internationales qui proposent la possibilité d'avorter.*

La différence avec l'exemple (38) d'*Élaboration* provient du fait que les expressions référentielles «Republican presidents» et «George W. Bush» ne renvoient pas au même référent, mais que «George W. Bush» est un élément de l'ensemble donné par «Republican presidents». Les événements décrits dans les segments (a) et (b) sont également dans une relation d'inclusion : «signed a ban on contributing money...» est un événement parmi l'ensemble des événements décrits dans «put limits». Comme nous l'avons souligné dans la section 2.2.3.1, l'*Exemplification* nous semble être un cas particulier d'*Élaboration*. Nous verrons par la suite que ces deux types d'exemple sont traités comme des élaborations dans le cadre de la SDRT.

2.3 Élaboration en génération automatique de texte

Les relations de discours ont joué un rôle important dans le domaine de la génération de texte (incluant la génération de dialogue, les systèmes de question-réponse...) (McKeown, 1985; Hovy, 1988a,b, 1989; Moore, 1989; Moore et Swartout, 1990; Paris, 1990; Cawsey, 1990; Maybury, 1990). Les auteurs s'appuient sur des théories du discours existantes. Hovy (1988a,b, 1989), Moore (1989), Moore et Swartout (1990) et Paris (1990) ont formalisé les relations de discours de la RST tandis que McKeown (1985) s'est appuyée sur les prédicats rhétoriques de Grimes (1975).

Nous souhaitons présenter dans cette section la ou les relations de discours autour de la notion d'élaboration qui ont été retenues dans le cadre de la génération de texte, Cf. Tableau 2.2.

Le tableau 2.2 est basé sur les relations de la RST, mais une relation supplémentaire est utilisée dans deux théories, nommée *Object-Function*. Les applications de génération de texte qui font appel aux relations de discours nécessitent des relations spécifiques. De ce fait, aucune théorie n'utilise la relation d'*Élaboration* telle quelle, mais plutôt des sous-types de celle-ci (en bas de la hiérarchie de Hovy et Maier (1991)).

Les applications de génération de texte ont également souligné le fait que l'utilisation de marques de surface permet d'améliorer la cohésion des textes produits. Les théories du

(Hovy et Maier, 1991)					
(McKeown, 1985)					
(Hovy, 1988a,b, 1989)					
(Moore, 1989; Moore et Swartout, 1990; Paris, 1990)					
(Cawsey, 1990)					
(Maybury, 1990)					
Élaboration					
Objet-attribut					
Objet-fonction					
Ensemble-membre					
Procès-étape					
Partie-tout					
Général-spécifique					
Abstrait-exemple					

TAB. 2.2 – Bilan des types d'Élaboration employés en génération automatique de textes

discours auxquelles elles font appel doivent être capable de mettre en corrélation relations de discours et marques de surface. Tout au long de ce chapitre, nous avons relevé des marques de discours qui peuvent signaler la relation ou la rendre explicite.

Relations	Marqueurs
Exemplification	for instance, for example, in particular, i.e., an example would be, etc.
Reformulation	or (rather), or in other words, a better way of putting it, it would be better to say, or again, namely, that is to say etc.
Résumé	in brief, in all, I will sum up by saying, I shall conclude by saying, etc.

TAB. 2.3 – Type de marqueurs de discours pour des sous-types de la relation d’*Élaboration*

Nous voyons à partir du tableau 2.3 que les marqueurs de discours que nous avons relevés ne couvrent pas tous les sous-types de la relation d’*Élaboration* et nous reviendrons sur ce point dans les chapitres qui suivent et notamment dans le chapitre 4 consacré à la question de la signalisation des relations de discours.

2.4 Synthèse

Les travaux que nous avons présentés dans ce chapitre montrent la nécessité de tenir compte de relations de discours intersegments (entre propositions, phrases, paragraphes...) pour guider les inférences et les processus d’organisation des textes. Nous avons montré que la notion d’*élaboration* est une notion incontournable, à la fois dans les théories du discours mais également dans des applications telles que la compréhension et la génération de texte. Enfin, nous avons montré la diversité des définitions et les difficultés rencontrées par les théories pour établir une définition robuste et rigoureuse.

La première section de ce chapitre a mis en exergue la variété des notions que l’on peut rapprocher de l’*élaboration*. De ce fait, il n’est pas étonnant que les théories du discours aient proposé une relation d’*Élaboration* dominant plusieurs sous-types (Mann et Thompson, 1987; Hovy et Maier, 1991; Halliday, 1985). Le tableau 2.4 propose un récapitulatif des sous-types d’*Élaboration* pris en compte dans les théories que nous avons présentées dans cette section sur la base de la taxonomie de Hovy et Maier (1991).

Certaines théories emploient une seule relation pour exprimer plusieurs concepts d’*élaboration*, par exemple la définition de Halliday (1985). D’autres théories, quant à elles, préfèrent distinguer plusieurs relations, par exemple Mann et Thompson (1987) distinguent l’*Élaboration* et la *Reformulation* et Hobbs (1990) et Kehler (2002) distinguent l’*Élaboration* et l’*Exemplification*. Nous partageons le point de vue de Hovy et Maier (1991) selon lequel il peut être suffisant de considérer une seule relation de discours pour exprimer le concept d’*élaboration* et que celle-ci peut être affinée selon les besoins et les buts de l’analyse. Nous serions d’avis d’inclure l’*Exemplification* comme un cas particulier d’*Elaboration*. De

(Hovy et Maier, 1991)							
(Williams, 1893)							
(Quirk et Greenbaum, 1973)							
(Grimes, 1975)							
(Reichman, 1978)							
(Halliday, 1985; Thompson, 2004)							
(Hobbs, 1979, 1985, 1990; Kehler, 2002)							
(Mann et Thompson, 1987, 1988)							
Élaboration d'objet							
Objet-attribut							
Objet-fonction							
Ensemble-membre							
Élaboration de partie							
Procès-étape							
Partie-tout							
Élaboration de généralité							
Général-spécifique							
Abstrait-exemple							
Autres							
Identification							
Reformulation							

TAB. 2.4 – Bilan des types d'*Élaboration* présentes dans quelques théories du discours

même pour la relation de *Reformulation*, nous considérons cette relation comme un cas particulier de l'*Élaboration* : toute reformulation apporte un changement de point de vue et apporte de ce fait une ou des informations supplémentaires. Une relation que nous avons évoquée dans la section 2.2.2.1 n'est pas présente dans le tableau 2.4. Il s'agit de la relation *Commentaire*. Cette relation relie deux segments lorsque le second donne le point de vue d'un personnage ou de l'auteur sur ce qui est décrit dans le premier. Il s'agit non pas d'une élaboration mais d'une évaluation d'un fait, d'un personnage ou d'un événement par un personnage ou par l'auteur. Enfin, nous rejetons la relation d'*Identification* présente dans le tableau 2.4 car celle-ci n'est pas une relation de discours mais un moyen de maintenir la cohésion (à travers les expressions référentielles) dans un texte.

Une des difficultés pour établir les définitions de relations de discours dans une théorie est liée au fait que les relations de discours prennent leur sens au sein d'une théorie où celles-ci se trouvent naturellement délimitées par d'autres relations de discours. Il faut donc que chaque relation de discours soit rigoureusement définie afin d'éviter les confusions entre celles-ci. Dans ce chapitre, nous avons déjà présenté des confusions possibles entre la relation d'*Élaboration* et les relations d'*Explication* et d'*Arrière-Plan*. Nous reviendrons sur ces confusions dans la suite de la thèse, dans le chapitre 9.

Kleiber et Vassiliadiou (2009) ont présenté les difficultés rencontrées par les théories pour définir la relation d'*Élaboration* comme résultant d'une tension entre le pôle de l'«identité» et le pôle de la «différence». Une bonne définition de la relation d'*Élaboration* doit, selon eux, rendre compte de cette tension.

Dans ce chapitre, nous avons opposé deux traditions : les théories fonctionnelles et les théories issues de l'Intelligence Artificielle. Ces deux types de théories ont été conçus dans des buts différents.

Les théories fonctionnelles mettent en avant des objectifs descriptifs. Ils proposent un inventaire des relations de discours qui leur paraissent nécessaires pour fournir l'analyse d'un texte et sont également sensibles au fait que cet inventaire peut varier selon le type de textes étudiés.

Les théories en Intelligence Artificielle, quant à elles, cherchent à donner les moyens pour un système informatique de reproduire les inférences produites par les êtres humains lors de l'interprétation d'un discours. L'accent est donc mis sur les processus cognitifs mis en oeuvre par les lecteurs pour interpréter un discours.

Nous allons poursuivre ce travail en présentant les définitions qui ont été successivement données à la relation d'*Élaboration* dans le cadre de la SDRT. La SDRT cherche à donner des règles pour rendre compte de l'interprétation d'un discours par le biais de sa représentation et de sa construction. Nous allons voir comment ces objectifs sont atteints avec l'évolution des définitions et/ou descriptions de la relation d'*Élaboration*.

Chapitre 3

La relation d'*Élaboration* en SDRT

Sommaire

3.1 Définitions	72
3.1.1 Définitions (Asher, 1993)	72
3.1.2 Calcul des relations temporelles (Lascarides et Asher, 1993) . . .	78
3.1.3 Descriptions (Asher et Lascarides, 2003)	79
3.1.3.1 Descriptions et propriétés d' <i>Élaboration</i>	79
3.1.3.2 Descriptions et propriétés de <i>Continuation</i>	82
3.1.3.3 Illustration	82
3.2 Inférer <i>Élaboration</i>	85
3.2.1 Le prédicat <i>Subtype_D</i>	85
3.2.2 Les cadres de discours	87
3.3 Une <i>Élaboration</i>, des <i>Élaborations</i> ?	89
3.3.1 Distinguer <i>Élaboration d'Entité</i> et <i>Élaboration d'Éventualité</i> . . .	89
3.3.1.1 Premières observations	89
3.3.1.2 Formalisation d'une nouvelle relation de discours, l' <i>Élaboration d'Entité</i>	89
3.3.2 Distinguer <i>Élaboration d'état</i> et <i>Élaboration d'événement</i>	92
3.3.3 Distinguer <i>Particularisation</i> et <i>Généralisation</i>	93
3.4 Synthèse	95

Après avoir exposé les définitions et les descriptions de la relation d'*Élaboration* au sein de nombreuses théories dans le chapitre 2, nous allons consacrer ce chapitre à celles qui ont été données successivement à la relation d'*Élaboration* dans la SDRT (Asher, 1993; Lascarides et Asher, 1993; Asher et Lascarides, 2003). Comme nous l'avons déjà dit, la SDRT est le cadre dans lequel se situent nos analyses. Cette théorie a un nombre restreint et fermé de relations de discours et en conséquence, la relation d'*Élaboration* dans cette théorie couvre un grand nombre de cas présentés dans le chapitre précédent.

Nous présentons, dans la section 3.1, les différentes versions et les évolutions des définitions et descriptions de la relation d'*Élaboration* dans le cadre de la SDRT. Présenter les évolutions des définitions de la relation nous permet d'évoquer, d'une part, des problématiques essentielles sur la façon de définir et/ou décrire les relations de discours dans

un cadre formel et d'aborder, d'autre part, des problématiques sur la définition de la relation d'Élaboration elle-même. Dans la section 3.2, nous présentons les différents axiomes permettant d'inférer la relation d'Élaboration. Puis en section 3.3, nous présentons des travaux proposant de nouvelles relations de discours, en quelque sorte dérivées de la relation d'Élaboration.

3.1 Définitions

3.1.1 Définitions (Asher, 1993)

Le livre de Asher (1993) fournit la première version de la SDRT et des définitions des relations de discours dans ce cadre. Dans cette version, la relation d'Élaboration est la seule relation subordonnante. Elle fait intervenir la notion de *domination discursive* et une relation orientée narrativement de *partie à tout* entre éventualités :

Def 1 *Elaboration* (α, β) iff $(\alpha \Downarrow \beta \vee (\text{for every } e \in ME(\beta) \text{ there is an } e' \in ME(\alpha) \text{ such that } e \text{ is a part of } e')) \ \& \ (\beta \text{ is more complex than } \alpha))$. (Asher, 1993, p. 267)¹

avec

- α est une SDRS (le constituant élaboré) ;
- β est une SDRS (le constituant élaborant) ;
- e est une éventualité (événement ou état) ;
- ME sont les éventualités principales (Main Eventualities) ;
- \Downarrow signifie domine discursivement (D-Domine) ;
- \vee est le symbole de la disjonction inclusive ;
- \in est le symbole de l'appartenance ;
- $\&$ est le symbole de la conjonction.

Cette définition exploite trois relations : la condition de *domination discursive* (\Downarrow), la condition de *partie à tout* entre éventualités et une condition de *complexité*. Deux conditions sont en disjonction inclusive : la condition de *domination discursive* et la conjonction de la condition de *partie à tout* avec la condition de *complexité*. La disjonction inclusive entraîne qu'une seule ou les deux conditions peuvent s'appliquer entre les constituants α et β . Nous exposons puis illustrons ces deux cas.

La condition de domination discursive ($\alpha \Downarrow \beta$) Cette condition repose sur une hiérarchie de types qui s'établit dans un ensemble de types. Cette hiérarchie de type couvre à la fois les types d'éventualités et d'objets. Elle fait partie d'une base commune de connaissances du monde. $\alpha \Downarrow \beta$ si et seulement si α subsume β ou si et seulement si pour tout x $\beta(x)$ implique $\alpha(x)$ mais pas l'inverse, i.e. $\alpha(x)$ n'implique pas $\beta(x)$.

¹Une parenthèse manquante dans le livre (Asher et Lascarides, 2003) a été rectifiée en accord avec N. Asher (communication personnelle).

Conjonction de la condition de partie à tout et de la condition de complexité ((pour chaque $e \in ME(\beta)$ il existe un $e' \in ME(\alpha)$ tel que e est une partie de e') & (β est plus complexe que α)) Cette condition exprime une relation de partie à tout (*part of*), i.e. une relation de sous-éventualité (*subevent*) entre les éventualités : «in which one event is part of or a subevent of another event» (Asher, 1993, p. 267) et une relation de complexité entre les constituants visant à rendre compte du fait que le β est plus complexe que le constituant α , autrement dit que le constituant β contient plus de détails, qu'il est plus riche, plus détaillé et plus précis du point de vue informationnel que le constituant α ². La condition de partie à tout se distribue sur un ensemble d'éventualités principales afin de permettre son application sur des éventualités plurielles, comme illustré en (1).

- (1) [L'été de cette année là vit plusieurs changements dans la vie de nos héros]_a. [François épousa Adèle]_b. [Jean-Louis partit pour le Brésil]_c et [Paul s'acheta une maison à la campagne]_d.

Les éventualités décrites en (b), (c) et (d) constituent chacune une partie de l'éventualité plurielle décrite en (a).

Dans (Asher, 1993), la relation d'*Élaboration* est présentée comme une relation complémentaire de la relation de *Continuation* car cette dernière relie deux constituants α et β seulement si ceux-ci partagent un topique commun. La relation de *Continuation* est une relation faible et compatible avec plusieurs relations de discours qui n'introduisent pas une domination discursive. La relation de *Continuation* relie deux constituants si le second prolonge la fonction rhétorique du premier. Pour exprimer cette notion, la SDRT emploie des constituants appelés *topiques*. Les topiques contribuent à la structure globale de la SDRS en fournissant des résumés à des groupes de plusieurs constituants (regroupés en un constituant complexe), conférant ainsi au constituant complexe dominé un rôle structurel particulier. Les structures topicales permettent d'organiser le contenu du discours. Les topiques sont toujours des constituants simples, i.e. avec un seul événement principal. Enfin, lorsque deux constituants α et β constituent une SDRS \mathcal{K} , α est un topique de β (soit $\alpha \Downarrow \beta$ dans \mathcal{K}) si et seulement si α est simple. Dans ce cas α est un topique explicite. De plus, dans cette version, \Downarrow implique *Élaboration*.

Passons maintenant à un autre type d'exemple. Chaque proposition dans (2) introduit un segment de discours minimal (DRS). Nous allons appliquer cette définition à l'analyse de l'attachement des DRS minimales de (2) dont la construction a été présentée dans le chapitre 1 section 1.3.1.2.

- (2) (a) Catherine mesure 157 cm. (b) Elle a les cheveux bruns et les yeux bleus. (c) Elle enseigne à l'université. (traduit de Asher (1993, p. 276-277))

Nous allons construire étape par étape la SDRS de (2).

Le constituant créé à partir du segment (a) s'insère dans un contexte discursif vide (étant donné que la DRS \mathcal{K}_a construit à partir de (a) est le premier constituant minimal de ce discours).

²Une définition formelle de la condition de complexité est fournie dans (Asher, 1993, p. 300).

Une fois les deux DRS construites à partir des segments (a) et (b), comment peut-on les relier l'une à l'autre dans une SDRS cohérente? Selon l'analyse proposée dans (Asher, 1993), la présence d'une relation anaphorique entre les deux DRS et l'utilisation dans les deux constituants du même temps verbal - le présent - sont des indices favorables - mais pas infaillibles - de la relation de *Continuation*. De plus, aucune autre relation de discours incompatible avec la relation de *Continuation* ne peut être inférée entre les deux constituants. Par conséquent, Asher (1993) infère que les deux constituants³ sont reliés par la relation de *Continuation*. Cet attachement est non seulement essentiel pour garantir la cohérence du discours mais également pour résoudre l'anaphore dans (b). Les deux constituants étant liés par la relation de *Continuation*, cela implique qu'ils partagent un topique commun selon la définition de la relation de *Continuation* dans cette version de la SDRT.

Nous construisons donc un constituant, la SDRS \mathcal{K} obtenue par le rattachement des constituants \mathcal{K}_a et \mathcal{K}_b par la relation de *Continuation*. Comme stipulé dans la définition de *Continuation*, ces deux constituants partagent un topique commun qui résume l'ensemble de la SDRS \mathcal{K} . Asher (1993) propose le topique «Catherine», nommé \mathcal{K}_0 . Selon la définition de *Topique*, si on a *Topique*(α, β) ou $\alpha \Downarrow \beta$, alors on a β est une élaboration de α .

La troisième et dernière phrase, dont la représentation est donnée dans la DRS \mathcal{K}_c , poursuit le même topique. Asher (1993) explique que trois points d'attachement sont possibles, \mathcal{K}_0 , \mathcal{K} et \mathcal{K}_b mais que la théorie ne nous dit pas lequel choisir. Une lecture attentive de l'exemple permet d'attacher le nouveau constituant \mathcal{K}_c à \mathcal{K} . La DRS \mathcal{K}_c construite à partir de (c) poursuit la description du topique «Catherine». La DRS \mathcal{K}_c se rattachant au constituant \mathcal{K}_b et poursuivant la description du topique *Catherine* est intégrée au constituant complexe \mathcal{K} , donnant la SDRS finale illustrée Figure 3.1.

Avec l'exemple (2), nous avons illustré la notion de construction des constituants topiques, qui sont liés à l'inférence de la relation d'*Élaboration* via la domination discursive. Nous allons maintenant illustrer la condition de *partie à tout* avec le prochain exemple (Asher, 1993, p. 279) :

- (3) (a) I ate a lovely dinner. (b) I had quenelles de brochet. (c) I had salmon. (d) I had duck. (e) I had a nice wine. (f) I then went for a walk around the old city.
 (a) J'ai mangé un agréable dîner. (b) J'ai eu des quenelles de brochet. (c) J'ai eu du saumon. (d) J'ai eu du canard. (e) J'ai eu du bon vin. (f) Je suis allé ensuite me promener dans la vieille ville.

Nous n'allons pas entrer dans le détail de la construction de chaque DRS construite à partir des segments de (3) pour nous concentrer sur les attachements des constituants *via* les relations de discours. La première phrase (a) introduit un constituant \mathcal{K}_a . La seconde phrase (b) introduit un constituant \mathcal{K}_b , qui doit être attaché à \mathcal{K}_a lors de la procédure de construction de la SDRS. Selon la règle de pertinence (*Rule of Relevance*) dans la SDRT, il faut effectivement déterminer une relation qui s'établit entre \mathcal{K}_a et \mathcal{K}_b étant donné que la succession des deux segments est supposée cohérente. Des connaissances du monde nous amènent à conclure de façon défaisable que l'événement principal de \mathcal{K}_b est une partie

³Dans la version de 1993, les relations de discours relient directement les constituants.

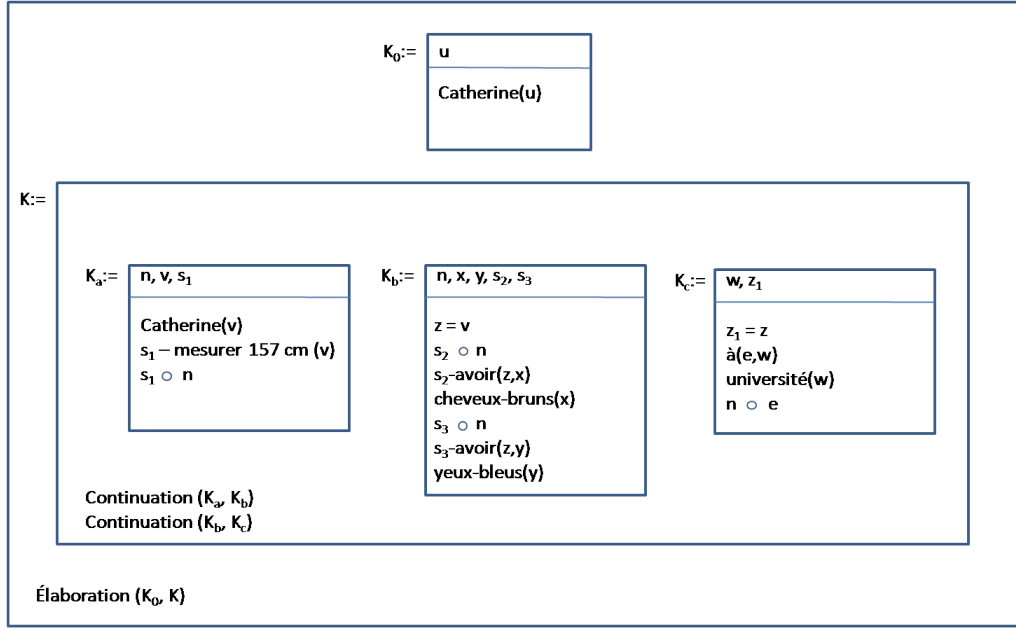


FIG. 3.1 – SDRS de (2)

de l'événement principal de K_a . La définition de la relation permet d'inférer *Élaboration* lorsque l'événement principal de β est une partie de l'événement principal de α . Il est alors possible de conclure que β est une élaboration de α : «manger des quenelles de brochet» est une étape du «repas». Ainsi, étant donné qu'aucune autre information ne vient contredire ce choix de relation, nous pouvons conclure que K_b est une élaboration de K_a . La phrase (c) introduit un constituant K_c . Deux attachements sont possibles pour ce constituant, K_a ou K_b . K_c poursuit la description du repas amorcée dans le constituant K_b . Il faut donc aussi vérifier que K_c est une élaboration de K_a . En effet, si K_b et K_c sont reliés par *Continuation*, la sémantique de cette relation implique que K_c poursuive la même fonction rhétorique que K_b . En poursuivant le même raisonnement que pour K_a et K_b on infère *Élaboration* entre K_a et K_c : «manger du saumon» est une seconde étape du «repas». Etant donné qu'*Élaboration*(K_a, K_b) et *Élaboration*(K_a, K_c) ont été inférées, l'attachement de K_b et K_c par *Continuation* est confirmé. Ces deux constituants K_b et K_c sont alors regroupés dans le segment complexe K qui a pour topique K_a . Des conclusions similaires sont apportées pour le rattachement de K_d et K_e , respectivement les constituants introduits par (d) et (e), par la relation de *Continuation* d'une part, et par la relation d'*Élaboration* d'autre part.

En ce qui concerne le rattachement du constituant K_f introduit par la phrase (f) : «I then went for a walk around the old city», trois attachements sont possibles K_a , K_e et K . L'éventualité décrite dans le constituant K_f , la «promenade», ne poursuit pas le même topique - «le repas» - que les constituants inclus dans le constituant complexe K . Par conséquent, K_f n'est pas une élaboration de K_a et K_f n'est pas une continuation de K_e . Selon l'analyse de (Asher, 1993), aucune information ne nous permet d'inférer une

quelconque relation entre \mathcal{K}_e et \mathcal{K}_f ou \mathcal{K}_e et \mathcal{K} . Enfin, (Asher, 1993) considère l'attachement de \mathcal{K}_f à \mathcal{K}_a : aucune information ne permet de définir la relation de discours qui s'applique entre ces deux constituants donc la relation choisie par défaut est celle de *Continuation*. Nous remarquons que le rôle du marqueur *then* n'est pas pris en compte dans l'analyse de cet exemple. Cette première version présentée dans (Asher, 1993) est une ébauche de la SDRT. Ce même exemple dans (Asher et Lascarides, 2003) prend en compte le marqueur *then*, permettant d'inférer la relation de *Narration* entre ces deux constituants.

La règle de construction du topique nous amène alors à construire un topique \mathcal{K}_0 qui subsume le constituant complexe \mathcal{K}_A formé de \mathcal{K}_a et \mathcal{K}_f , dont le contenu est «la soirée».

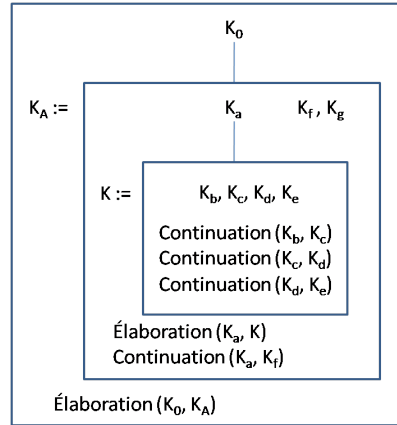


FIG. 3.2 – SDRS de (3)

Bilan La définition de la relation d'*Élaboration* chez (Asher, 1993) recouvre trois configurations différentes :

- La condition de domination discursive ($\alpha \Downarrow \beta$), illustrée avec les constituants topiques (2) ;
- La conjonction de la condition de partie à tout (for every $e \in \text{ME}(\beta)$ there is an $e' \in \text{ME}(\alpha)$ such that e is a part of e') et de la condition de complexité, illustrée avec (3) ;
- Les deux conditions réunies.

Condition de domination discursive Contrairement à Hobbs (1979, 1985, 1990), Asher (1993) définit la relation d'*Élaboration* comme une relation subordonnante. Il note que la relation d'*Élaboration* «does not introduce just a two-place relation, but rather a complex condition in which one constituent *discourse dominates* or *d-dominates* another - much in the same way that a more specialized topic in a taxonomy of some domain.» (Asher, 1993, p. 266). Asher (1993), contrairement à Hobbs (1979, 1990) qui met en avant le pôle de l'identité, met au premier plan la différence entre les deux constituants à travers la hiérarchie de type pour la première condition et la condition de complexité pour la seconde. Toutefois, la définition de la relation de domination discursive dans (Asher, 1993) rend également compte du pôle de l'identité : " $\alpha \Downarrow \beta$ means that α is a semantic label for

β ; it holds just in case α acts as a summary or topic statement of β ." (Asher, 1993, p. 267). Les deux constituants sont décrits comme véhiculant un même contenu.

L'exemple suivant, extrait du corpus ILF⁴ et analysé dans Kleiber et Vassiliadiou (2009, p. 196), illustre cette condition de domination discursive :

- (4) [Au petit matin, hier, vers 5h30, un homme de 28 ans, passablement éméché, a importuné des policiers qui procédaient à un contrôle routier sur les allées Jean-Jaurès à Toulouse.],_a [Celui-ci a menacé les fonctionnaires avec sa cannette de bière,],_b [tout en émaillant ses propos de quelques noms d'oiseaux.],_c (*La Dépêche du Midi*, 09/07/03)

Il est effectivement possible d'établir une hiérarchie entre les types des éventualités présentes ici. Les types d'événements 'X menacer Y avec sa cannette de bière' et 'X émailler ses propos de quelques noms d'oiseaux' sont des sous-types de 'X importuner Y'. De plus, on s'assure ici que tous les X et tous les Y renvoient respectivement aux mêmes référents. On peut également rendre compte de l'implication entre \mathcal{K}_a et \mathcal{K}_b , et entre \mathcal{K}_a et \mathcal{K}_c tel que décrit dans la définition de la domination discursive :

- *X menace Y avec sa cannette de bière* \rightarrow *X importune Y*
- **mais** *X importune Y* \nrightarrow *X menace Y avec sa cannette de bière*
- *X émaille ses propos de quelques noms d'oiseaux* \rightarrow *X importune Y*
- **mais** *X importune Y* \nrightarrow *X émaille ses propos de quelques noms d'oiseaux*.

Condition de partie à tout Cette condition est, comme le souligne Asher (1993), «more narrative». Elle stipule qu'une éventualité est une partie ou une sous-éventualité d'une autre éventualité. Nous allons analyser ici un exemple similaire (5) déjà traité dans le chapitre 2.

- (5) [Mixel a escaladé le Vignemale hier matin.],_a [Il est parti du refuge des Oulettes au lever du jour,],_b [puis il a passé la Hourquette d'Ossoue vers 9h,],_c [et il est arrivé au sommet vers midi.],_d ((Bras, 2007, p. 232))

Dans cet exemple, les trois phrases (b) (c) et (d) élaborent la première (a). Chacune des phrases (b), (c) et (d) dénote un événement qui est une partie ou un sous-événement de l'événement rapporté dans la première phrase (a) : «partir du refuge», «passer la Hourquette d'Ossoue» et «arriver au sommet» sont trois étapes de l'événement «escalader le Vignemale».

La condition de domination discursive ne s'applique pas directement à cet exemple car les événements «X partir du refuge des Houlettes» et «X passer la Hourquette d'Ossoue vers 9h» ne sont pas des sous-types de «X escalader le Vignemale». De même, les constituants \mathcal{K}_b et \mathcal{K}_c n'impliquent pas \mathcal{K}_a :

- *X partir du refuge des Houlettes* \nrightarrow *X escalader le Vignemale*
- *X passer la Hourquette d'Ossoue vers 9h* \nrightarrow *X escalader le Vignemale*

En revanche, il est possible d'appliquer la conjonction de la condition de partie à tout et de la condition de complexité : les événements principaux de \mathcal{K}_b , \mathcal{K}_c et \mathcal{K}_d sont des parties

⁴Corpus constitué par Francis Cornish pour le projet ILF «Relations de Cohérence et Fonctionnement des Anaphores.

ou sous-événements de l'événement principal de \mathcal{K}_a et chacun de ces trois constituants contient plus de détails sur l'événement décrit dans le constituant \mathcal{K}_a .

Les deux conditions réunies Comme la disjonction inclusive l'indique, une ou l'autre ou les deux conditions peuvent intervenir dans l'inférence de la relation d'*Élaboration*. L'exemple (1) fait intervenir les deux conditions⁵.

(1) [L'été de cette année là vit plusieurs changements dans la vie de nos héros]_a. [François épousa Adèle]_b. [Jean-Louis partit pour le Brésil]_c et [Paul s'acheta une maison à la campagne]_d.

En (1) où l'éventualité décrite en (a) est une éventualité plurielle, on a en effet une relation de partie à tout, les éventualités décrites en (b), (c) et (d) sont des parties des «changements dans la vie de nos héros». Dans le même temps, chacune des éventualités décrites en (b), (c) et (d) est un type d'éventualité qui peut être un sous-type de «change-ment», la relation de domination discursive est donc également établie.

La condition de domination discursive se base sur une hiérarchie de types entre objets et éventualités, proche de la notion d'hypéronymie en linguistique. Elle est calculée hors contexte selon les types d'objet et d'éventualité posés dans le discours. La condition de partie à tout, quant à elle, se base sur des éventualités en contexte. La relation de partie à tout ne porte pas sur les types mais sur les éventualités elles-mêmes. Il s'agit donc d'une relation mise en place dans le discours. Nous verrons avec la description de la relation dans (Asher et Lascarides, 2003) que ces deux niveaux mériteraient d'être clairement distingués.

D'un point de vue formel, les deux parties de la définition de la relation d'*Élaboration* dans (Asher, 1993) sont dans une relation de bi-implication («iff» ou «ssi», «si et seulement si»). Dans la bi-implication, chacune des parties de la définition implique l'autre. Autrement dit, d'une part, la vérification d'une des conditions de domination discursive et/ou de partie à tout au niveau sémantique implique la relation d'*Élaboration* au niveau discursif et de l'autre, la présence de la relation d'*Élaboration* au niveau discursif implique la vérification d'une des conditions de domination discursive et/ou de partie à tout au niveau sémantique. Cette bi-implication permet d'avoir une définition complètement spécifiée mais un effet, non voulu, de cette définition pose problème : la séparation entre niveau sémantique et niveau rhétorique (ou discursif) passe au travers de cette définition (Vieu, 2007).

3.1.2 Calcul des relations temporelles (Lascarides et Asher, 1993)

L'article de Lascarides et Asher (1993) apporte des informations complémentaires sur la caractérisation de certaines relations de discours et plus particulièrement sur le calcul des relations temporelles entre les événements introduits par les constituants que les relations de discours relient.

Les définitions en «ssi» sont abandonnées au profit d'axiomes exprimant les effets sémantiques d'une relation. La description d'une relation de discours est ainsi obtenue par

⁵Cet exemple constitue un contre-exemple aux arguments avancés par Kleiber et Vassiliadiou (2009) selon lesquels il n'est pas possible d'avoir un exemple faisant intervenir les deux conditions réunies.

la combinaison de deux règles, la première pour l'inférence de la relation de discours et la seconde pour le calcul des relations temporelles ou autres effets sémantiques et discursifs de cette dernière.

Considérons les nouveautés apportées à la définition de la relation d'*Élaboration*. Lascarides et Asher (1993) décrivent la relation de façon informelle comme celle qui s'établit entre deux constituants α et β si l'événement de β est une partie de l'événement α (peut-être en étant sa phase préparatoire) tel que dans l'exemple ci-dessous :

- (6) [The council built the bridge.]_a [The architect drew up the plans.]_b (Lascarides et Asher, 1993, p. 438)

La description de la relation d'*Élaboration* est obtenue par la combinaison de deux règles : la première *défaisable* $>$ (pour l'inférence de la relation de discours) et la seconde *indéfaisable* \rightarrow pour le calcul des effets sémantiques de la relation de discours.

Pour inférer la structure de ce mini-discours, la règle suivante (Lascarides et Asher, 1993, p. 465) s'applique :

Axiome 3.1 Inférer Élaboration Phase Préparatoire

$\langle \tau, \alpha, \beta \rangle \wedge prep(me(\beta), me(\alpha)) > \text{Élaboration}(\alpha, \beta)$

Cette règle spécifie que, dans une structure où l'on considère le rattachement du constituant β au constituant α , si l'éventualité principale du constituant β est une phase préparatoire de celui du constituant α , alors on peut inférer de façon défaisable que les deux constituants sont reliés par la relation d'*Élaboration*.

Enfin, un axiome (Lascarides et Asher, 1993, p. 465) donne de façon indéfaisable une conséquence de l'inférence de cette relation de discours entre deux constituants :

Axiome 3.2 Effets discursifs ($\text{Élaboration}(\alpha, \beta) \rightarrow \neg \text{Narration}(\alpha, \beta)$)

Si deux constituants sont reliés par la relation d'*Élaboration*, alors on peut inférer de façon indéfaisable que la relation de *Narration* ne peut pas être vérifiée entre ces deux constituants. Par la suite, cet axiome sera une conséquence des effets sémantiques de base (la précédence temporelle impliquée par *Narration* est incompatible avec l'inclusion temporelle impliquée par *Élaboration*).

Dans la prochaine section, nous allons présenter la dernière version de la description de la relation d'*Élaboration* dans le cadre de la SDRT.

3.1.3 Descriptions (Asher et Lascarides, 2003)

3.1.3.1 Descriptions et propriétés d'*Élaboration*

La présentation de la relation d'*Élaboration* dans la SDRT (Asher et Lascarides, 2003) reprend un grand nombre d'éléments présents dans la version de 1993 : la notion de subordination et la notion de domination discursive se retrouvent dans la définition du prédicat de subsumption présent dans la règle de déclenchement de la relation ; et la condition de partie à tout se retrouve dans les effets sémantiques de la relation.

Les relations de discours dans Asher et Lascarides (2003) sont associées d'une part à des *règles de déclenchement* qui permettent de les inférer et d'autre part à des *effets sémantiques*. Pour la relation d'Élaboration, nous avons donc :

Une règle de déclenchement La relation d'Élaboration peut être inférée de façon non monotone (ou défaisable) grâce à la présence d'une relation de subsumption (Asher et Vieu, 1995) qui existe entre les types des éventualités des segments à relier. Cette relation est exprimée par le prédicat $Subtype_D$ ⁶ qui indique que le type de la seconde éventualité est un sous-type de la première en s'appuyant sur la sémantique lexicale des éventualités considérées et/ou sur des connaissances du monde relatives au contexte discursif.

Axiome 3.3 Inférer Élaboration

$$(\text{?}(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge Subtype_D(\sigma, \beta, \alpha) \wedge Aspect(\alpha, \beta) > \acute{E}laboration(\alpha, \beta, \lambda) \\ (Asher \text{ et } Lascarides, 2003, p. 206))$$

avec

- $\text{?}(\alpha, \beta, \lambda)$ qui signifie que l'attachement des constituants α et β doit être déterminé.
- $Top(\sigma, \alpha)$ qui signifie que le constituant α est compris dans un constituant complexe σ , et que celui-ci n'est compris dans aucun autre constituant.
- $Aspect(\alpha, \beta)$ qui signifie que les constituants α et β sont caractérisés par la même valeur aspectuelle (état ou événement)

Des effets sémantiques Les effets sémantiques de la relation d'Élaboration sur le contenu du discours font intervenir une relation de partie à tout, telle que définie dans (Aurnague et Vieu, 1993), notée *part-of*. Autrement dit, l'éventualité élaborante est une partie (ou une partie propre selon les versions de la SDRT) de l'éventualité élaborée. Cette relation de partie à tout implique à son tour une inclusion temporelle entre les éventualités, c'est à dire que la temporalité de l'éventualité du segment élaborant est incluse (strictement ou au sens large selon les versions de la SDRT) dans celle de l'éventualité élaborée.

Les effets spatio-temporels de l'Élaboration sont décrits en terme d'une inclusion stricte dans (Asher *et al.*, 1995, p. 32) :

Axiome 3.4 Effets Sémantiques (1^{ère} version)

$$\phi_{\acute{E}laboration\alpha,\beta} \rightarrow ProperPart - of(e_\beta, e_\alpha) \\ ProperPart - of(e_\alpha, e_\beta) \rightarrow e_\beta \subset e_\alpha$$

Les effets spatio-temporels sont étendus de l'inclusion stricte à l'inclusion (large) dans (Asher et Lascarides, 2003, p. 160) afin d'inclure notamment les cas de reformulation d'une même éventualité :

Axiome 3.5 Effets Sémantiques (2^{ème} version)

$$\phi_{\acute{E}laboration\alpha,\beta} \rightarrow Part - of(e_\beta, e_\alpha) \\ Part - of(e_\alpha, e_\beta) \rightarrow e_\beta \subseteq e_\alpha$$

⁶Sous-type dans le discours donné (D).

Enfin, la relation d'*Élaboration* implique la relation dite *Topique* et notée \Downarrow dans la version de 2003.

Axiome 3.6 *Élaboration implique* \Downarrow

$$Elaboration(\alpha, \beta) \Rightarrow \Downarrow (\alpha, \beta)$$

(Asher et Lascarides, 2003, p. 227)

Cet axiome spécifie que si les deux constituants α et β sont reliés par la relation d'*Élaboration* alors α est un constituant topique pour β .

Seule la relation de *Topique* peut être inférée. Dans cette version, la relation *Topique* n'implique plus *Élaboration*. Elle entraîne une différence majeure qui est la mise à jour du constituant topique avec les informations sémantiques contenues dans les constituants dominés.

Asher et Lascarides (2003) décrivent ensuite des propriétés qui sont liées à la relation d'*Élaboration*.

La propriété de transitivité La relation d'*Élaboration* est transitive. Autrement dit, si deux constituants π_α et π_β sont reliés par la relation d'*Élaboration* et deux constituants π_β et π_χ sont également reliés par la relation d'*Élaboration*, alors on peut inférer la relation d'*Élaboration* entre π_α et π_χ :

Axiome 3.7 *Règle de transitivité* (illustrée Figure 3.3)

$$(Elaboration(\pi_\alpha, \pi_\beta) \wedge Elaboration(\pi_\beta, \pi_\chi)) \rightarrow Elaboration(\pi_\alpha, \pi_\chi)$$

(Asher et Lascarides, 2003, p. 161)

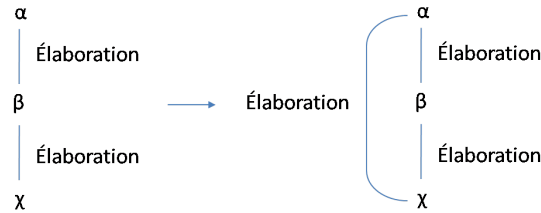


FIG. 3.3 – Application du principe de la transitivité

La propriété de distributivité Celle-ci est un principe de cohérence sur les structures du discours. Cette propriété spécifie que des attachements à des constituants subordonnés ont la même fonction discursive par rapport aux constituants superordonnés. Ce principe stipule que les éléments constitutifs d'un segment β , élaborant α , sont également des parties de ce segment α . En quelque sorte, les effets sémantiques de la relation d'*Élaboration* se distribuent sur toutes les éléments constituant le premier niveau du segment élaborant (qu'ils soient reliés par une relation coordonnante ou subordonnante). De ce fait, on peut attacher au moyen d'une relation coordonnante ou d'une relation subordonnante un constituant γ à un autre constituant β qui est lui même attaché à α via la relation d'*Élaboration*, seulement si ce constituant γ peut lui-même être rattaché à α via la même relation de discours.

L'analyse des discours en SDRT fait également intervenir l'*inverse* du principe de distributivité. Si on a $Elaboration(\alpha, \beta)$ et $Elaboration(\alpha, \gamma)$ alors β et γ sont reliés par la relation de *Continuation*.

3.1.3.2 Descriptions et propriétés de *Continuation*

La relation de *Continuation* est une relation coordonnante qui n'implique pas d'effets temporels. La règle de déclenchement de cette relation est basée sur l'environnement discursif : quand deux constituants sont reliés par une même relation subordonnante au même constituant dominant, alors la relation de *Continuation* est inférée entre ces deux constituants (Prévot, 2004) :

Axiome 3.8 Inférer *Continuation*

$$(?(\beta, \delta, \lambda) \wedge Elaboration(\alpha, \beta, \lambda) \wedge Elaboration(\alpha, \delta, \lambda)) > Continuation(\beta, \delta, \lambda)$$

3.1.3.3 Illustration

Nous allons maintenant illustrer l'application de ces principes en présentant l'analyse proposée par Asher et Lascarides (2003, p. 223) de l'attachement par des relations de discours des constituants obtenus à partir des segments de l'exemple suivant :

- (7) [Max experienced a lovely evening last night.]_a [He had a fantastic meal.]_a [He ate salmon.]_c [He devoured lot of cheese.]_d [He won a dancing competition.]_e
 [Max a eu une agréable soirée hier soir.]_a [Il a eu un repas fantastique.]_b [Il a mangé du saumon.]_c [Il a dévoré une grande quantité de fromage.]_d [Il a gagné une compétition de danse.]_e

On obtient les DRS \mathcal{K}_{π_a} et \mathcal{K}_{π_b} ⁷ à partir respectivement des segments (a) et (b). A ce niveau de la construction de la SDRS, un seul point d'attachement, π_a , est possible pour le constituant π_b . Nous devons donc mettre à jour la SDRS avec $?(\pi_a, \pi_b, \pi_0)$, π_0 étant la SDRS \mathcal{K}_{π_0} globale qui contient \mathcal{K}_{π_a} et \mathcal{K}_{π_b} .

La proposition (b) contient un pronom *he*, ce qui signifie que la SDRS de π_b contient une condition sous-spécifiée $x = ?$. Selon les principes du MDC, il existe une co-dépendance entre l'inférence des relations de discours et la résolution anaphorique. Dans le cas où un constituant contient un pronom, il faut que son attachement par une relation de discours permette sa résolution, i.e. que le constituant cible de l'attachement contienne un référent étant un antécédent possible. Il est possible d'inférer une relation de discours particulière entre π_a et π_b et selon les contraintes d'accessibilité des référents, le seul antécédent possible pour x est le référent de discours m introduit par le nom propre «Max» dans la SDRS \mathcal{K}_{π_a} . Donc $x = m$. Sur la base de connaissances du monde à propos de ce que les gens font durant leur soirée, par exemple «prendre un repas», il est possible d'inférer $Subtype_D(\pi_b, \pi_a)$. La résolution du pronom «he», au préalable, indique que les agents des deux événements sont co-référents, ce qui est, par ailleurs nécessaire pour inférer $Subtype_D$, qui nécessite que les agents des deux événements soient le même. De cette façon, $Elaboration(\pi_a, \pi_b)$ est inférée

⁷Les relations de discours, dans (Asher et Lascarides, 2003), relient les étiquettes π des constituants \mathcal{K} , Cf. Chapitre 1.

de façon non monotone (défaisable). Les effets sémantiques de la relation doivent maintenant être vérifiés entre les événements des constituants \mathcal{K}_{π_a} et \mathcal{K}_{π_b} : l'événement «having a fantastic meal» est, comme prédit, temporellement inclus dans l'événement «having a lovely evening».

À présent, trois sites d'attachement sont possibles pour $\pi_c : \pi_b$ (qui est la dernière information nouvelle), π_a (qui domine π_b) et π_0 (qui est le noeud le plus haut dans la structure). L'attachement de π_c à π_0 n'est pas satisfaisant car il n'y a pas d'antécédents accessibles dans la SDRS de \mathcal{K}_0 pour résoudre l'anaphore *he* de (c). Sur la base de connaissances du monde, on peut inférer *Subtype_D* entre un événement du type «eating salmon» et un événement du type «enjoying an evening». La relation entre les constituants π_a et π_c est donc *Élaboration*. Mais dans ce cas, nous pourrions également inférer une connection faible (*Continuation*) entre π_b et π_c : en d'autres termes, cela représente une situation étrange dans laquelle «Max eating salmon» et «his having dinner» sont deux événements reliés par une relation faible, peut-être deux dîners. L'attachement de π_c à π_b est alors considéré. La SDRS de \mathcal{K}_{π_b} contient un antécédent pour la résolution de l'anaphore et sur la base de connaissances du monde, on peut inférer *Subtype_D* entre un événement du type «eating salmon» et un événement du type «having meal». On peut donc inférer *Élaboration*(π_b, π_c). Selon les principes du MDC, on choisit la SDRS dont l'interprétation offre une cohésion plus forte entre la nouvelle information «eating salmon» et les anciennes informations «eating a meal», «having a lovely evening». Avec l'attachement *Élaboration*(π_b, π_c), on peut également inférer *Élaboration*(π_a, π_c) en appliquant la règle de transitivité. Les connections sont, en quelque sorte, maximisées. Le principe du MDC prédit que \mathcal{K}_c s'attache à \mathcal{K}_b pour garantir une cohérence du discours maximale.

Quatre sites d'attachement sont disponibles pour $\pi_d : \pi_c$ (qui est la dernière information nouvelle), π_b (qui domine π_c), π_a (qui domine π_b) et π_0 . Il est possible d'inférer *Narration*(π_c, π_d) sur la base du prédicat *Occasion* (inférée sur la base de connaissances du monde à propos du déroulement d'un repas et de l'ordre dans lequel les plats sont servis). *Narration* entraîne une contrainte de topique. On vérifie donc si une SDRS peut jouer ce rôle. C'est le cas de π_b . Par ailleurs, π_b est déjà le topique du constituant π_c . Selon le principe de la distributivité, tout constituant attaché par une relation coordonante à un constituant qui fait partie d'une *Élaboration* fait lui-même partie de cette *Élaboration*. Donc π_b est le topique du constituant complexe π_B , formé des constituants π_c et π_d . Cette solution engendre de plus une cohésion plus forte du discours que si l'on assume que le topique est non mentionné dans le discours.

Le principe du MDC joue également un rôle dans l'attachement du constituant π_e . Les sites d'attachement possibles sont les suivants : π_0 , π_a , π_b , π_B et π_d (le constituant π_c ne se trouve plus sur la frontière droite). Pour garantir le principe du MDC, le choix du rattachement de π_e doit être celui qui propose le maximum de connexion cohésive dans le discours. Considérons les sites d'attachement possibles. Il n'y a pas d'axiomes qui permettent d'inférer *Occasion*, *Cause* ou *Subtype* pour le rattachement de π_e à π_d . Les mêmes conclusions sont obtenues pour le rattachement π_e à π_B . Le rattachement de π_e à π_a entraîne un discours plus cohérent. Il est possible d'exploiter le prédicat *Subtype_D* pour attacher π_e à π_a , «winning a dancing competition» est un type d'événement qui peut être un sous-type du type de l'événement «having a lovely evening». Donc on infère *Élaboration*(π_a, π_e). De ce fait, «eating the meal» et «winning a dancing competition» sont

tous les deux des événements qui font partie de la soirée. Ces informations sont vérifiées par l'axiome *Occasion* entre π_b et π_e . Dans ce cas, étant donné que nous avons déjà inféré $Elaboration(\pi_a, \pi_b)$ et $Narration(\pi_b, \pi_e)$, en vertu du principe *inverse* de la distributivité, π_a peut être considéré comme un constituant complexe topique de la séquence narrative constituée de π_b et π_e , étiqueté π_A . L'attachement de π_e est un exemple de *Discourse Pop*, c'est à dire qu'il n'est pas attaché au constituant précédant mais à un constituant introduit plus tôt dans la structure discursive. Ce *Discourse Pop* est le résultat d'une mise à jour de la SDRS qui maximise la cohérence.

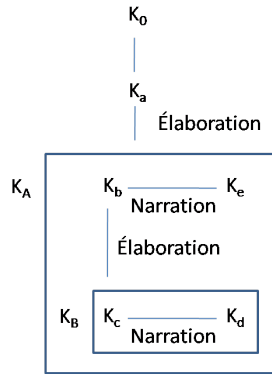


FIG. 3.4 – SDRS de (7)

Enfin les constituants π_b , π_B , π_c et π_d ne sont plus désormais des sites d'attachement possibles car ils ne sont plus sur la frontière droite. Cela permet de prédire que la poursuite en (8) produit un discours inapproprié :

- (8) It was a beautiful pink.
Il était rose

puisque pour résoudre l'anaphore *it* qui renvoie au *saumon*, il faudrait attacher ce constituant à K_c , qui ne se trouve plus sur la frontière droite et se trouve donc indisponible pour un rattachement. Cette contrainte prédit également que la poursuite en (9) produit un discours inapproprié :

- (9) He then had a soufflé.
Il a ensuite eu un soufflé

car cet événement dénote une partie du repas mais l'attachement à K_b ou à K_d n'est plus autorisé.

L'analyse de (7) fait intervenir la relation de *Subtype_D* entre les types d'événements «have a fantastic meal» et «eat salmon». La même relation est, en (5), inférée entre les types d'événements «escalader le Vignemale» et «partir du refuge des Oulettes» dans les analyses de Bras (2007).

En (5) repris ci-dessous, (Bras, 2007) propose l'analyse suivante de l'attachement des constituants par des relations de discours :

- (10) (5) [Mixel a escaladé le Vignemale hier matin.]_a [Il est parti du refuge des Oulettes au lever du jour]_b, [puis il a passé la Hourquette d'Ossoue vers 9h.]_c [et il est arrivé au sommet vers midi.]_d

Le constituant π_a décrit un événement de type «escalader le Vignemale» et les constituants π_b , π_c et π_d décrivent des événements qui sont des étapes de l'événement décrit en π_a . Nous inférons, en premier lieu, *Subtype_D* entre «escalader le Vignemale» et «partir du refuge des Oulettes», grâce au contenu sémantique des verbes «escalader» et «partir de» et à des connaissances du monde relatives aux randonnées dans les Pyrénées. Donc, nous inférons *Élaboration*(π_a, π_b). Nous inférons ensuite la relation de Narration entre π_b et π_c . Cette relation est déclenchée ici par le marqueur *puis*. Le constituant π_c est donc relié, au moyen d'une relation coordonnante, au constituant π_b , formant le constituant complexe, π_A , lui-même relié au constituant π_a par *Élaboration*. Dans ce cas, le principe de la distributivité stipule que le constituant π_c est aussi relié au constituant π_a par *Élaboration*. Le connecteur *et* et les informations temporelles contenues dans π_b et π_c nous permettent d'inférer Narration entre π_c et π_d , et d'appliquer le même principe que précédemment pour relier π_d à π_a par *Élaboration*. Le constituant complexe π_A est constitué des constituants simples π_b , π_c et π_d . Et, π_A élabore π_a .

Nous présentons ci-après, plus en détail, ce prédicat *Subtype_D* employé dans les analyses de (7) et (5).

3.2 Inférer *Élaboration*

3.2.1 Le prédicat *Subtype_D*

Les informations contenues dans le prédicat *Subtype_D* s'appuient essentiellement sur une hiérarchie de type établie au niveau lexical. Les informations contenues dans la hiérarchie de type permettent d'inférer *Élaboration* au niveau du discours *via* le prédicat *Subtype_D*.

Le lexique inclut des informations à propos des types sémantiques des objets dénotés par les noms communs, les verbes, les adjectifs, *etc.* Un sous-type est relié à un super-type par la notion de substituabilité. Le sous-type hérite des caractéristiques du super-type et se distingue par des caractéristiques spécifiques. Le sous-type peut être substitué par le super-type mais l'inverse n'est pas nécessairement possible. Le concept de *Subtype_D* est proche de la notion linguistique d'hyperonymie.

Cependant, les informations contenues dans la hiérarchie de type ne suffisent pas à elles seules pour inférer le prédicat *Subtype_D* qui nécessite que d'autres types d'informations lexicales soient codées dans le lexique, tel qu'illustré dans l'exemple suivant par Asher et Lascarides (2003, p. 282) :

- (11) [Max ate a lovely meal.]_a [He devoured lots of salmon.]_b

Dans l'exemple (11), les phrases (a) et (b) contiennent des items lexicaux liés sémantiquement. Premièrement, le type de l'événement dénoté en e_a , «to eat» est un super-type de

celui dénoté en e_b , «to devour». D'un point de vue linguistique, ces deux verbes sont reliés par la relation sémantique de troponymie⁸.

L'analyse de Asher et Lascarides (2003) tient également compte du lien sémantique qui relie les arguments des verbes des deux phrases (a) et (b), i.e. «meal» et «salmon». Cependant dans la hiérarchie de type, «salmon» n'est pas directement un sous-type de «meal». D'autres informations lexicales sont nécessaires pour inférer ce lien sémantique entre les deux items lexicaux. D'abord, dans la hiérarchie de type «salmon» est un sous-type de «animal» et est un nom comptable. A partir de cette entrée dans la hiérarchie de type est automatiquement créée une entrée lexicale pour le sens de nourriture de «salmon» qui est un nom de masse. Ensuite, le mot «meal» dénote un événement qui implique l'action de manger «eating the meal». Cette propriété est également celle des noms de masse (au sens de nourriture) dérivés des noms comptables d'animaux. En définitive les mots «salmon» et «meal» sont ainsi rapprochés par la propriété commune qu'ils partagent de «pouvoir être mangé». Mais cela ne permet pas de définir lequel des deux est un sous-type de l'autre.

Asher et Lascarides (2003) prennent x et y (référents du discours respectivement pour «meal» et «salmon»), représentés tel que suit $y \sqsubseteq x$, dans la *glue logic*, pour spécifier que dans le lexique le type assigné à y est un sous-type du type assigné à x . L'événement dénoté en e_a de «eating» et l'événement dénoté en e_b de «devouring» sont contraints de façon similaire, de sorte que le type assigné à e_a est un sous-type de celui assigné à e_b . Enfin, «meal» et «salmon» ont le même rôle thématique de patient.

Les deux propositions (a) et (b) contiennent donc des éléments qui se trouvent dans une relation de sous-type. Ces informations sont utilisées au niveau du discours pour inférer *Subtype_D* entre les constituants α et β . En d'autres termes, dans les propositions α et β , les éléments «salmon» et «meal» ont le même rôle thématique, noté respectivement $\theta_i(x, \alpha)$ et $\theta_i(y, \beta)$, avec $i = \text{patient}$. Ces deux arguments sont dans une relations de sous-type, notée $y \sqsubseteq x$. Enfin le type de l'événement principal de β est un sous-type du type de l'événement principal de α , noté $e_\beta \sqsubseteq e_\alpha$. On peut ainsi inférer la relation de sous-type entre les constituants α et β :

Axiome 3.9 Inférer *Subtype_D* $(\theta_i(x, \alpha) \wedge \theta_i(y, \beta) \wedge y \sqsubseteq x \wedge e_\beta \sqsubseteq e_\alpha) \rightarrow \text{Subtype}_D(\beta, \alpha)$
(Asher et Lascarides, 2003, p. 283)

Nous venons d'introduire le prédicat *Subtype_D* tel qu'il est présenté et expliqué dans (Asher et Lascarides, 2003).

Nous voyons dans les analyses des exemples (5), (7) et (11) que ce prédicat est utilisé avec une acception plus large pour tenir également compte des cas d'*Élaboration* dans lesquels l'éventualité élaborante est une étape de l'éventualité élaborée, ce que Kleiber et Vassiliadiou (2009) ont mis en évidence. Nous proposons de définir deux prédicats différents, sur lesquels s'appuieraient deux règles de déclenchement pour *Élaboration* : un prédicat *Subtype_D* qui encode des informations de sous-type et un prédicat *Subevent_D* qui encode des informations sur la structure d'une éventualité en sous-éventualités. Ces deux prédicats encoderaient ainsi deux types d'informations pré-existantes au discours.

⁸La troponymie est une relation sémantique s'établissant entre deux verbes de modification par la manière.

Axiome 3.10 Inférer Élaboration Subevent

$$(?(\alpha, \beta, \lambda) \wedge \text{Top}(\sigma, \alpha) \wedge \text{Subevent}_D(\sigma, \beta, \alpha) \wedge \text{Aspect}(\alpha, \beta) > \text{Élaboration}(\alpha, \beta, \lambda))$$

Les compléments phrastiques subordonnés Asher et Lascarides (2003) proposent un axiome supplémentaire pour inférer la relation d'Élaboration basé sur une construction syntaxique dans laquelle se manifeste nécessairement la relation de sous-type. Asher et Lascarides (2003, p. 285) proposent l'analyse de l'exemple suivant dans lequel les types des événements principaux des phrases (a) et (b), respectivement «to made a promise» et «to phone» sont, par dérivation, dans une relation de sous-type, un événement de type «to promise to phone» étant un sous-type de l'événement de type «to promise something» :

- (12) [John made a promise to Mary.]_a [He would phone her.]_b (Asher et Lascarides, 2003, p. 285)
 [John fit une promesse à Marie.]_a [Il lui téléphonerait.]_b

La même relation discursive présente entre ces deux phrases peut être obtenue en utilisant une autre structure syntaxique que la simple consécution comme dans les deux exemples suivants :

- (13) [John made a promise to Mary.]_a [that he would phone her.]_b (Asher et Lascarides, 2003, p. 285)
 [John fit une promesse à Marie.]_a [qu'il lui téléphonerait.]_b
- (14) [John promise to Mary.]_a [that he would phone her.]_b (Asher et Lascarides, 2003, p. 285)
 [John promet à Marie.]_a [qu'il lui téléphonerait.]_b

L'axiome, appelé *Post Head Sentential Complements*, exploite directement une information de type syntaxique pour l'inférence de la relation d'Élaboration :

Axiome 3.11 Post Head Sentential Complements

$$(?(\alpha, \beta, \lambda) \wedge [\text{that}(x, p)](\alpha) \wedge p = \beta) \rightarrow \text{Élaboration}(\alpha, \beta, \lambda) \quad (3.1)$$

(Asher et Lascarides, 2003, p. 285)

Si α décrit un objet x (ici «promise») qui prend un complément phrastique optionnel introduit par *that* – p (qui introduit le constituant β), la proposition p est sous-spécifiée dans α mais résolue dans β ($p = \beta$), alors on infère de façon monotone, au niveau discursif, que β est une élaboration de α .

Dans la section suivante, nous allons présenter comment l'hypothèse de l'encadrement temporel du discours (Charolles, 1997) a été traduite dans le cadre de la SDRT en faisant appel à la relation d'Élaboration.

3.2.2 Les cadres de discours

Vieu *et al.* (2005) ont cherché à exprimer l'hypothèse de l'encadrement temporel du discours (Charolles, 1997) dans le cadre de la SDRT. Selon Charolles (1997), les adverbiaux

temporels détachés à gauche⁹ introduisent un référent temporel qui sert d'index pour l'interprétation d'une ou de plusieurs phrases.

- (15) **Hier au petit matin**, un homme de 28 ans a importuné des policiers qui procédaient à un contrôle routier sur les allées de Jean-Jaurès à Toulouse. Celui-ci a menacé les fonctionnaires avec sa cannette de bière, tout en émaillant ses propos de quelques noms d'oiseaux. (adapté de *La Dépêche du Midi*, 09/07/03)

Dans l'exemple (15), l'adverbe de localisation temporelle «Hier au petit matin» introduit un cadre temporel qui indexe toutes les propositions de l'exemple sur la base d'un critère commun : chaque proposition du cadre est indexée par l'information temporelle donnée par l'adverbial ou introducteur de cadre, «Hier au petit matin».

Vieu *et al.* (2005) ont montré, qu'en plus du rôle d'indexation de l'adverbial temporel sur les phrases qui sont dans sa portée, les introducteurs de cadre signalent l'intention du locuteur/scripteur de localiser une suite d'événements par rapport à un moment précis, spécifié par l'adverbial temporel détaché à gauche, comme si celui-ci annonçait pour l'exemple (15) : «je vais te raconter quelque chose qui s'est passé *hier au petit matin*». De plus, les introducteurs de cadre sont considérés comme des segments de discours qui permettent d'introduire dans la structure du discours un constituant dominant plusieurs constituants simples regroupés dans un constituant complexe. Vieu *et al.* (2005) traitent donc les introducteurs de cadre comme des *Introducteurs de Nouveau Topique* (INT). Le topique, introduit par l'introducteur de cadre, domine par la suite tous les constituants obtenus à partir des phrases qui se trouvent dans la portée de l'introducteur de cadre. Ces topiques sont sous-spécifiés au moment de leur introduction dans la structure du discours et leur contenu se trouve ensuite déterminé par le ou les constituants à venir dominés par le topique. Le rôle de l'INT est plus fort que le rôle de l'adverbial de localisation temporelle étant donné qu'il introduit dans la représentation du discours un événement sous-spécifié. En plus d'indexer les propositions sous sa portée comme ayant eu lieu «hier au petit matin», il introduit dans le discours un événement complexe sous spécifié, «ce qui s'est passé hier au petit matin». Enfin, la relation de discours qui relie le topique au constituant complexe dominé est la relation d'*Élaboration* étant donné que les propositions qui se trouvent sous sa domination viennent spécifier cet événement complexe.

Dans les sections suivantes, nous allons présenter des évolutions récentes concernant la relation d'*Élaboration* dans le cadre de la SDRT. Chacune des sous-sections ci-dessous va présenter une distinction binaire permettant de catégoriser avec plus de finesse la relation d'*Élaboration*. D'abord, nous présenterons les travaux de Prévot *et al.* (2009) qui distinguent l'*Élaboration d'événement* et l'*Élaboration d'entité*; ensuite nous aborderons des problèmes concernant les effets sémantiques de l'*Élaboration d'événement* vs. de l'*Élaboration d'état*; enfin, nous présenterons les travaux de Danlos et Gaiffe (2000) qui distinguent, au lieu d'une unique *Élaboration*, deux relations de discours *Particularisation* et *Généralisation*.

⁹Dans sa thèse, Ho-Dac (2007) montre que la contrainte du détachement à gauche n'est pas suffisante pour accorder le rôle d'introducteur de cadre à un adverbial temporel. Elle montre qu'il faut tenir compte de la position dans le texte de cet adverbial détaché à gauche (début de paragraphe, début de section...).

3.3 Une *Élaboration*, des *Élaborations* ?

3.3.1 Distinguer *Élaboration d'Entité* et *Élaboration d'Éventualité*

3.3.1.1 Premières observations

Dans le chapitre 2, nous avons présenté un état de l'art des définitions données à la relation d'*Élaboration* mettant notamment en avant le fait que l'élément élaboré peut être soit l'éventualité principale dénotée par le segment ou une ou des entités introduites dans le segment. Généralement ces deux types d'*Élaboration* sont inclus dans une même définition (Mann et Thompson, 1987; Hovy et Maier, 1991).

Fabricius-Hansen et Behrens (2001) sont les premières, à notre connaissance, à avoir dégagé cette distinction entre ce qu'elles nomment *e[ventuality]-Elaboration*, i.e. élaboration sur des référents du discours (éventualités) établis par des phrases ou des propositions et *i[n]dividual-Elaboration*, i.e. élaboration ou description de référents du discours (objets ou participants), sans toutefois exprimer des définitions formelles ou descriptives de ces deux types.

Knott *et al.* (2001) posent le problème de la relation *Object/Attribute Elaboration* de la RST. Selon Mann et Thompson (1987), cette relation est vérifiée entre deux segments de texte si le noyau contient la mention d'un *objet* et si le satellite présente des *attributs* de cet objet. Knott *et al.* (2001) interprètent comme *attributs* toutes les propositions qui donnent des informations supplémentaires sur l'*objet* en question. La principale critique faite par Knott *et al.* (2001) concerne le fait qu'effectivement la relation *Object/Attribute Elaboration* n'est pas vérifiée entre des propositions (alors que c'est le cas pour les autres relations de la RST) mais entre une proposition et une entité contenue dans la proposition cible de la relation. De plus, ils notent qu'il n'existe pas de marqueurs du discours pour cette relation en anglais (ce qui constitue une entrave à l'hypothèse de Knott (1996), *Cf.* . (15)). Ces raisons ont amené Knott *et al.* (2001) à ne pas considérer cette relation comme faisant partie de l'ensemble des relations de discours. Leur solution consiste à considérer une structure globale du texte basée sur les continuations et les changements de topiques d'entité. L'usage des relations de discours est alors restreint à la modélisation de la cohérence locale.

Selon Prévot *et al.* (2009), si ce traitement peut fonctionner sur des textes ou des passages visant la description d'un objet, d'un personnage, d'un paysage, etc., il ne correspond pas à la structure des textes ou des passages narratifs et argumentatifs pour lesquels la structure globale est régie par des relations de discours et pour lesquels ce n'est que localement que certaines entités sont caractérisées par un contenu informationnel supplémentaire. Ils vont donc proposer de distinguer, dans le cadre de la SDRT, une nouvelle relation de discours, l'*Élaboration d'Entité*.

3.3.1.2 Formalisation d'une nouvelle relation de discours, l'*Élaboration d'Entité*

Prévot *et al.* (2009) partent d'un constat présent dans de nombreux travaux d'annotation à grande échelle de relations discursives (Reese *et al.*, 2007; Wolf et Gibson, 2005) : la confusion dans l'annotation de deux relations de discours, la relation d'*Élaboration* et la relation d'*Arrière-plan*. Nous avons mis en évidence, dans le chapitre précédent, le manque de consensus général et de précision concernant la relation d'*Élaboration*. Ceci se trouve

confirmé par les difficultés rencontrées par les annotateurs lors d'une annotation systématique sur des corpus de textes attestés. En partant de ce constat sur la confusion des deux relations de discours, *Élaboration* et *Arrière-plan* (Wolf et Gibson, 2005; Reese *et al.*, 2007), Prévot *et al.* (2009) cherchent à affiner l'établissement des règles d'identification des deux relations. Cela les amène à introduire une nouvelle relation de discours, l'*Élaboration d'Entité*.

L'exemple suivant illustre la confusion entre la relation d'*Élaboration* et la relation d'*Arrière-plan* (Prévot *et al.*, 2009, p. 208) :

- (16) [Pourquoi a-t-on abattu Paul Mariani, [cinquante-cinq ans]_a, [attaché au cabinet de M. François Doubin]_b, [ministre délégué]_c, [chargé du commerce et de l'artisanat]_d?]_e (Le Monde)

Les segments (a), (b), (c) et (d) apportent des informations additionnelles. Ils peuvent être interprétés comme des éléments secondaires, un arrière-plan pour le segment (e) ou comme des propriétés décrivant un objet du discours, ici *Mariani* du segment (e) pour les segments (a) et (b) et «François Doubin» du segment (b) pour les segments (c) et (d).

Les auteurs posent d'abord la question de la nécessité de considérer ces segments (a), (b), (c) et (d) (ici des appositions) comme des segments de discours. La question de la nature des segments du discours est une question plus générale que l'on ne va aborder que partiellement dans cette section (pour plus de détails, voir section 8.2.2.1 du chapitre 9 pour les consignes de segmentation appliquées sur le corpus ANNODIS). Prévot *et al.* (2009) évoquent effectivement la solution de ne pas les considérer comme des segments de discours et de ne considérer qu'un seul segment pour la totalité de cette phrase, proposant ainsi une approche plus traditionnelle de l'interface syntaxe/discours. Ils refusent cette solution en notant que celle-ci, plutôt que de simplifier la tâche de l'annotateur, rend plus difficile la reconnaissance de la fonction discursive du segment global. Ils préfèrent résoudre l'apparente ambiguïté du rattachement de ces petits segments ((a), (b), (c) et (d)) et proposer ainsi une annotation plus fine. Le même choix a été retenu dans le cadre du projet ANNODIS : toutes les appositions sont segmentées.

Les relations de discours en SDRT (1.3.2.3 dans le chapitre sur les structures et relations de discours) sont caractérisées de façon stricte et précise. Pour rappel :

- La relation d'*Élaboration* est inférée sur la base de la relation qui s'établit entre les types des éventualités décrites par les segments sur la base de connaissances lexicales et/ou de connaissances du monde. Plus précisément, si le type de l'éventualité e_β introduite par le second segment est un sous-type du type de l'éventualité e_α introduite par le premier segment, nous pouvons inférer *Elaboration*(α, β). De plus une contrainte aspectuelle pèse sur les éventualités reliées, il doit s'agir soit de deux états, soit de deux événements.
- La relation d'*Arrière-plan* est inférée sur la base d'un critère aspectuel. Une relation d'*Arrière-plan* est inférée si les segments présentent une nature aspectuelle différente, i.e. un segment décrit un état tandis que le segment le succédant ou le précédant fait référence à un événement. L'état décrit toujours l'arrière-plan.

Du point de vue des règles de déclenchement de ces deux relations, nous voyons donc que là où une *Élaboration* se vérifie, on ne peut pas vérifier un *Arrière-plan* et inversement.

Bien que les relations de discours soient établies sur la base de définitions solides dans le cadre de la théorie, leur utilisation par des annotateurs non spécialistes sur des corpus s'est révélée problématique. La confusion entre la relation d'*Élaboration* et la relation d'*Arrière-plan* peut être liée à plusieurs difficultés (Prévot *et al.*, 2009) :

- Premièrement, il n'est pas toujours facile de déterminer l'aspect (état/événement) des segments à relier. Cela est encore plus vrai pour les appositions (généralement sans verbes).
- Deuxièmement, au niveau théorique, les effets temporels de ces relations ne sont pas mutuellement exclusifs, autrement dit l'effet temporel de l'*Élaboration*, à savoir l'inclusion temporelle, est un cas particulier de l'effet temporel de l'*Arrière-plan*, à savoir le recouvrement. Donc du point de vue temporel, toutes les *Élaborations* pourraient *a priori* être des *Arrière-plans*.

Cette confusion peut également être tout simplement liée au fait qu'il manque une relation de discours pour décrire avec exactitude les textes. Cette nouvelle relation se trouve alors révélée par la confrontation entre la théorie et des nouvelles données attestées.

La solution proposée par Prévot *et al.* (2009) va effectivement dans ce sens. Ils repèrent une forte confusion entre les deux relations en question et constatent que les deux relations sont pourtant bien définies dans le cadre de la SDRT. Ils tirent donc la conclusion qu'un certain nombre de cas ne sont ni des élaborations, ni des arrière-plans (ni une autre relation dans la théorie) et pourraient correspondre à une nouvelle relation de discours qu'ils proposent d'introduire dans le cadre de la SDRT : la relation d'*Élaboration d'Entité*. Cette relation correspond à la relation *Object-Attribut Elaboration* de la RST, examinée dans (Knott *et al.*, 2001) et correspond à l'*Individual-Elaboration* chez (Fabricius-Hansen et Behrens, 2001).

Dans l'exemple (16), les segments (a) et (b) décrivent des propriétés (des attributs au sens de la RST) de l'entité *Paul Mariani*. Ils dénotent des états, et plus particulièrement des *états descriptifs*¹⁰. Mais, d'une part, ces états ne font pas partie de l'événement dénoté dans le segment (e) («l'assassinat de Paul Mariani») donc la relation d'*Élaboration* est exclue ; et d'autre part, ces états ne décrivent pas le décor ou l'arrière plan de la situation décrite par l'événement en (e) («l'assassinat de Paul Mariani») donc la relation d'*Arrière-plan* est également exclue. Il en est de même pour les segments (c) et (d) qui décrivent des propriétés de l'entité *François Doubin*. Ce sont des états qui ne font pas partie de l'événement dénoté dans le segment (e) et ces états ne sont clairement pas des décors ou des arrière-plans de «l'assassinat de Paul Mariani», donc la même conclusion s'impose : les relations d'*Élaboration* et d'*Arrière-plan* sont exclues.

Les segments (a), (b), (c) et (d) sont des élaborations d'entité dont la description est donnée dans (Prévot *et al.*, 2009, p. 215). La relation d'*Élaboration d'Entité* relie deux

¹⁰Les *états descriptifs* exposent des propriétés caractéristiques d'une entité, i.e. des propriétés permanentes telles que «être intelligent», «avoir les yeux verts»... Les propriétés transitoires, i.e. accidentelles ou temporaires telles que «être debout», «être malade»... donnent lieu à d'autres relations de discours, dont principalement *Arrière-plan*.

segments α et β , tel que :

- (i) α dénote une proposition introduisant une éventualité principale e_α (événement ou état) et un référent de discours accessible x ; e_α peut être un état, mais ce n'est pas un état descriptif de $x(\neg Etat - Descriptif(e_\alpha, x))$
- (ii) β dénote une proposition dont l'éventualité principale est un état e_β descriptif de $x(Etat - Descriptif(e_\beta, x))$
- (iii) e_β et e_α se recouvrent temporellement ($O_t(e_\alpha, e_\beta)$)

La relation d'*Élaboration d'entité* est déclenchée par un changement de perspective : lorsque β est un état descriptif d'une entité x et que α n'est pas un état descriptif de x :

Axiome 3.12 $(?(\alpha, \beta) \wedge P(e_\alpha, \vec{y}) \wedge x \in \vec{y} \wedge \neg Indice - Etat - Descriptif(e_\alpha, x) \wedge Indice - Etat - Descriptif(e_\beta, x) \succ E - Elab(\alpha, \beta)$

La relation d'*Élaboration d'Entité* est une relation subordonnante. En SDRT, une continuation d'une relation subordonnante implique la création d'un constituant complexe regroupant les constituants qui poursuivent la même fonction rhétorique, comme nous l'avons vu pour la relation d'*Élaboration* (Cf. 3.1.3). Le même principe s'applique pour l'*Élaboration d'Entité* : pour l'exemple (16), les constituants π_a et π_b forment un constituant complexe, π_A . Ce constituant complexe est une *Élaboration d'Entité* du constituant π_e . De même les constituants π_c et π_d forment un constituant complexe, π_B . Ce constituant complexe est une *Élaboration d'Entité* du constituant π_b .

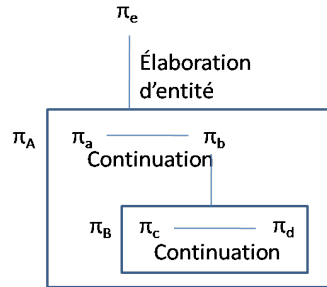


FIG. 3.5 – SDRS des segments (a) à (b) (16)

Il est cependant nécessaire de contraindre ce principe pour traiter les exemples où deux segments sont des élaborations de deux entités d'un même segment. Dans ce cas, Les auteurs proposent d'introduire un nouveau type de constituant implicite, *le topique d'entité* qui précise l'entité élaborée et bloque le regroupement de segments qui sont des élaborations d'entité qui ne portent pas sur la même entité.

3.3.2 Distinguer *Élaboration d'état* et *Élaboration d'événement*

Les exemples prototypiques de la relation d'*Élaboration* que nous avons présentés jusque là concernaient uniquement des événements. Mais comme nous l'avons déjà dit, la relation

d'Élaboration peut relier deux constituants qui décrivent soit deux événements, soit deux états. Nous allons donc à présent analyser des exemples d'Élaboration dans lesquels les constituants reliés sont deux états.

- (17) [Un soir,]_a [il faisait un temps horrible,]_b [les éclairs se croisaient,]_c [le tonnerre grondait,]_d [la pluie tombait à torrent ;]_e [c'était épouvantable !]_f [Quelqu'un frappa à la porte du château,]_g [et le vieux roi s'empessa d'ouvrir.]_h (Wikisource, La Princesse au petit pois)

Nous pouvons établir *Subtype_D* entre le type de l'état «il faisait un temps horrible» et les types de chaque état suivant : «les éclairs se croisaient», «le tonnerre grondait», «la pluie tombait à torrent» sur la base de connaissances du monde. En conséquence, la relation d'Élaboration s'applique entre le constituant π_b et les constituants π_c , π_d et π_e . π_c , π_d et π_e poursuivent la même fonction rhétorique, i.e. ce sont des élaborations du constituant π_b qui joue le rôle de topique dans la structure. Ces trois constituants sont donc reliés entre eux par la relation de *Continuation*. Comme nous l'avons vu, la relation de *Continuation* est une relation dont la sémantique est faible. De plus, nous avons vu que les effets sémantiques de l'Élaboration stipulent que l'éventualité élaborante est une partie de l'éventualité élaborée, ce qui a pour conséquence au niveau spatio-temporel que l'extension temporelle de l'éventualité élaborante est incluse dans celle de l'éventualité élaborée. Cet axiome, nous l'avons vu, fonctionne bien pour les cas d'Élaboration qui portent sur des événements. En revanche l'extension spatio-temporelle d'un état n'est pas aussi clairement délimitée que celle d'un événement. Les états, introduits par des phrases à l'imparfait, décrivent des situations en cours de déroulement par rapport à la narration des événements : «quelqu'un frappa à la porte», «le vieux roi s'empessa d'ouvrir». Ces cas rendent nécessaire que les effets sémantiques de l'Élaboration admettent une équivalence entre les extensions spatio-temporelles des éventualités en jeu dans la relation. Il est ainsi possible de traiter ces deux types d'élaboration avec la seule définition de la relation d'Élaboration de (Asher et Lascarides, 2003).

3.3.3 Distinguer *Particularisation* et *Généralisation*

Danlos et Gaiffe (2000) étudient la coréférence événementielle entre deux phrases et les relations de discours. Ils s'intéressent donc tout particulièrement à la relation d'Élaboration, en distinguant deux relations de discours *Particularisation* et *Généralisation*. Dans un article plus récent, Danlos (2004) propose de formaliser ces deux relations dans le cadre de la SDRT.

Danlos et Gaiffe (2000) étudient un cas particulier de coréférence événementielle. Ils s'intéressent à la configuration $\langle P_a, P_b \rangle$ où P_a et P_b sont deux phrases qui renvoient à la même action physique comme dans l'exemple (18) :

- (18) [Fred a sali un vêtement.]_a [Il a taché une chemise.]_b (Danlos et Gaiffe, 2000)

Pour décrire la coréférence événementielle, Danlos et Gaiffe (2000) font appel à deux types de relation de coréférence : la *particularisation* et la *généralisation*. Dans la configu-

ration de deux descriptions successives notées D_a et D_b d'une même entité notée x (pouvant être un événement ou un objet), D_a et D_b sont dans une relation de

- *particularisation* notée $D_b = \text{PART}(D_a)$ si et seulement si D_b apporte au moins une information nouvelle sur x par rapport aux informations données dans D_a ;
- *généralisation* notée $D_b = \text{GEN}(D_a)$ si et seulement si D_b n'apporte aucune information nouvelle sur x par rapport aux informations données dans D_a .

L'exemple (18) est un exemple de coréférence événementielle de type *particularisation*. Les deux GN indéfinis *un vêtement* et *une chemise* sont coréférents. Danlos et Gaiffe (2000) remarquent que ce sont des emplois inhabituels des indéfinis (qui ont toujours été décrits comme ayant une lecture existentielle, c'est-à-dire introduisant un nouveau référent de discours (Kamp et Reyle, 1993, entre autres). Il s'agit également d'un emploi inhabituel du couple hyperonymie/hyponymie qui est généralement introduit dans le discours dans le sens inverse : l'hyponyme apparaît en premier, puis il est repris anaphoriquement au moyen d'un hypéronyme.

La coréférence s'établit sur la base de connaissances lexicales : le verbe «salir» est un hyperonyme du verbe «tâcher» et de même le nom «vêtement» est un hyperonyme du nom «chemise».

La coréférence événementielle de type *généralisation* produit des discours dans lesquels la seconde phrase n'apporte pas d'information nouvelle mais est une reformulation présentant l'événement sous un jour nouveau, comme dans l'exemple suivant :

- (19) [Fred a tâché une chemise.]_a [Il a (donc) sali un vêtement.]_b (Danlos et Gaiffe, 2000)

La seconde phrase généralise la première, étant donné que «tâcher» et «chemise» dans la première phrase sont respectivement des hyponymes de «salir» et «vêtement» de la seconde phrase.

A partir de ses observations, Danlos propose donc l'hypothèse suivante :

Une relation de coréférence événementielle de type *particularisation* s'applique si et seulement si chaque élément de la seconde phrase est un hyponyme, une anaphore ou une répétition d'un élément correspondant (s'il existe) dans la première phrase .

Une relation de coréférence événementielle de type *généralisation* s'applique si et seulement si chaque élément dans la seconde phrase (comportant un connecteur comme *donc*) est un hyperonyme, une anaphore ou une répétition d'un élément correspondant dans la première phrase.

Dans un article plus récent, Danlos (2004) propose de passer de ces relations de coréférence événementielle aux relations de discours dans le cadre de la SDRT.

Concernant les types de *particularisation*, la SDRT, ainsi que d'autres théories de discours, proposerait la relation d'*Élaboration* pour traiter des exemples tels que (18). Toutefois Danlos postule l'existence d'une nouvelle relation de discours, appelée *Particularisation* qui implique une relation de coréférence événementielle. L'argument de Danlos est que, dans la SDRT et également dans les autres théories du discours, la définition de la relation d'*Élaboration* n'implique pas de coréférence événementielle. Toutefois dans la version la plus récente de la SDRT, le prédicat *Subtype_D* est vérifié entre les prédicats des deux phrases,

un événement est sous-type de lui-même et les conséquences spatio-temporelles sont également vérifiées, étant donné que *Part – of* inclut les cas de recouvrement spatio-temporel total. Nous considérons donc la relation de *Particularisation* comme un cas particulier d'*Élaboration*, qui peut effectivement être utile dans des tâches de compréhension de texte ou de génération de texte comme le signale Danlos (2004).

Concernant les types de *généralisation*, Danlos propose également une nouvelle relation de discours *Généralisation*, qui implique une relation de coréférence événementielle que l'on peut voir comme un cas particulier de *Reformulation*. Cette position tient à condition que l'on considère la *Reformulation* comme une relation de discours à part entière. Dans les exemples de *Généralisation* donnés par Danlos, nous aurions plutôt tendance à inférer la relation de *Résultat* sur la base du marqueur *donc* que Danlos et Gaiffe (2000) insèrent eux-même dans leurs exemples de *Généralisation*, Cf. (19).

Nous ne considérons donc que la relation de *Particularisation* comme un cas d'*Élaboration*.

3.4 Synthèse

Les relations disponibles concernant l'élaboration sont à l'heure actuelle, dans le cadre de la SDRT au nombre de deux : la relation d'*Élaboration* qui porte sur des éventualités et la relation d'*Élaboration d'Entité* qui porte sur des entités (relation qui couvre des cas très précis, reliant une entité à des propriétés permanentes). Elles permettent de recouvrir les élaborations présentées dans le tableau 3.1.

À ces relations, nous pouvons ajouter la relation de *Topique*, relation subordonnante sous-spécifiée qui implique une mise à jour du constituant topique avec le contenu sémantique des constituants dominés. Une fois la mise à jour du topique effectuée, cette relation est similaire à la relation d'*Élaboration*.

Les définitions des relations de discours en SDRT sont strictes et précises, fournissant ainsi un cadre d'analyse dans lequel les frontières entre relations sont bien établies.

Les points qui restent à améliorer concernant la relation d'*Élaboration* dans la SDRT et qui nous intéressent dans ce travail concernent de notre point de vue les axiomes d'inférence. Les règles de déclenchement sont des heuristiques qui mettent en évidence la façon dont un analyste reconnaît la relation de discours. Ces règles font intervenir des éléments lexicaux, grammaticaux, syntaxiques, pragmatiques...

Le prédicat *Subtype_D* s'applique entre des types d'éventualités dont l'un est un sous-type de l'autre. Nous avons déjà vu que la relation de sous-type est plus large qu'une simple hiérarchie de type et nécessite un lexique enrichi. Il serait utile d'avoir des indications plus précises sur les enrichissements nécessaires dans le lexique pour parvenir à l'inférence de la relation de sous-type (hyperonymie, méronymie, synonymie...). Nous avons mis en évidence la nécessité d'introduire un prédicat *Subevent_D* pour rendre compte des exemples où un événement *y* est un type d'événement qui est sous-événement de *x*.

Types d'Élaboration	Élaboration (SDRT)	Élaboration d'entité (SDRT)	Exemples
Élaboration d'objet			
Objet-attribut			(16)
Objet-fonction			
Ensemble-membre			
Élaboration de partie			
Procès-étape (Subevent)			(5)
Partie-tout			(1)
Élaboration de généralité			
Général-spécifique			
Subtype			(4), (1)
Abstrait-exemple			
Autres			
Identification			
Reformulation			(18)

TAB. 3.1 – Bilan des types d'Élaboration en SDRT

Il faut maintenant vérifier que ces deux prédicats suffisent pour rendre compte de toutes les élaborations. Nous souhaitons proposer de nouvelles inférences, à partir de marqueurs lexicaux et syntaxiques, qui viennent compléter celles déjà existantes pour la relation d'*Élaboration*. Nous verrons, au fur et à mesure des analyses, que les mêmes marqueurs peuvent signaler la relation d'*Élaboration* et la relation de *Topique*. Pour nos objectifs d'analyse d'une part et d'inventaire d'autre part, nous relèverons indifféremment ces marqueurs qu'ils soient de la relation d'*Élaboration* ou de la relation de *Topique*.

Du point de vue du repérage automatique, la question se pose de la disponibilité de ces informations lexicales dans une ressource. De plus, le cadre d'analyse de la SDRT fournit des inférences sur la base de connaissances lexicales et de connaissances du monde qui se restreignent à un discours particulier, posant à nouveau la question de la disponibilité de ces informations dans une quelconque ressource. Ce type d'analyse semble, à ce jour, uniquement faisable par des analystes humains. Une façon d'alimenter le cadre de la SDRT serait d'une part de trouver des ressources lexicales disponibles pour envisager un repérage automatique de la relation et d'autre part de se fier à des marqueurs de surface, plus facilement accessibles automatiquement. Le chapitre suivant va porter sur la signalisation des relations de discours en général, puis de la relation d'*Élaboration* en particulier.

Chapitre 4

La signalisation des relations de discours

Sommaire

4.1	Marqueurs du discours dans les théories du discours	100
4.1.1	Caractère labile des marqueurs du discours	100
4.1.2	Rôle des marqueurs du discours dans les théories du discours . .	104
4.1.3	Synthèse sur indices et marqueurs du discours	111
4.2	Autres indices de la signalisation discursive	112
4.2.1	Des indices moins étudiés	112
4.2.2	Corrélation entre indices et relations de discours	112
4.3	Rôle des indices de la signalisation discursive en TAL	113
4.3.1	La génération automatique de texte	114
4.3.2	Le résumé automatique de texte	115
4.4	La signalisation de la relation d'Élaboration	116
4.4.1	Une relation non marquée et une relation par défaut	116
4.4.1.1	Une relation non marquée	116
4.4.1.2	Une relation par défaut	117
4.4.2	Des études sur la signalisation de la relation d'Élaboration	118
4.4.2.1	SPRIK : le marqueur <i>indem</i>	118
4.4.2.2	Les <i>Second-level Discourse Markers</i> de Siepmann (2005)	119
4.4.2.3	LEXCONN : liste de marqueurs lexicaux	121
4.4.2.4	Le projet ILF : Relations de cohérence et fonctionne- ment des anaphores	121
4.5	Bilan et positionnement	122

Les sources d'information mises en oeuvre dans la reconnaissance des relations de discours sont très riches mais souvent peu explicitées. Dans ce chapitre, l'objectif est de dégager les moyens mis en oeuvre par le lecteur ou allocutaire pour inférer les relations de discours. Plus particulièrement, nous cherchons à appréhender la façon dont les lecteurs ou allocutaires reconnaissent une relation de discours à partir de l'interaction entre les éléments disponibles à la surface des textes et le contenu des éléments reliés. Notre hypothèse

de départ repose sur le fait qu'à l'écrit, le texte contient des indices qui vont permettre au lecteur d'interpréter des successions de segments :

A l'écrit, communication distanciée, le texte constitue indubitablement la majeure partie des indices sur la base desquels le lecteur va construire un modèle interprétatif, modèle changeant qui informe à tout moment l'interprétation du segment en cours de lecture, et qui en est lui-même transformé. (Péry-Woodley, 2000, p. 43)

De plus, l'objectif d'identification systématique de la relation d'*Élaboration* que nous nous sommes fixée requiert un inventaire aussi riche que possible des indices disponibles à la surface des textes pour le repérage automatique.

Les indices, dans une perspective d'identification et de repérage dans les textes, sont composés de tous les éléments pouvant intervenir dans la construction du modèle interprétatif.

Les marqueurs, dans une perspective d'analyse, sont composés de tous les éléments dont la ou les fonctions discursives ont été établies. Les marqueurs d'une relation du discours sont ceux que l'on retrouve dans les règles d'inférence de la SDRT.

La section 4.1 de ce chapitre propose un état de l'art sur les éléments du texte appelés *marqueurs du discours*. Ce sont les éléments qui ont été les plus étudiés en tant que *signaux* de la structure du discours. Leur appellation indique qu'ils ont une fonction au sein du discours, qu'ils «marquent» une fonction discursive. Nous allons essentiellement discuter leur caractère labile et les fonctions discursives que nous pouvons leur attribuer. La section 4.2 va s'intéresser aux autres types d'indices, moins étudiés, de la structuration discursive. Puis la section 4.3 présente le rôle des indices, identifiables à la surface des textes, dans des applications de TAL. Enfin, la section 4.4 aborde la question plus précise de la signalisation de la relation d'*Élaboration*.

4.1 Marqueurs du discours dans les théories du discours

Oates (1999) propose un état de l'art sur les marqueurs du discours où elle relève que ce terme couvre un ensemble hétérogène de mots ou d'expressions et qu'il n'existe pas de consensus général à travers les différentes études qui emploient ce terme. En plus du terme *discourse markers* (Schiffrin, 1987), d'autres expressions sont utilisées, comme par exemple *cue phrases* (Grosz et Sidner, 1986), *cue words* (Hovy, 1993), *discourse cues* (Di Eugenio *et al.*, 1997), *discourse connectives* (Webber *et al.*, 2000), *connectives* (Elhadad et McKeown, 1990), *rhetorical markers* (Scott et De Souza, 1990), *clue words* (Reichman, 1978). Ces termes diffèrent dans leur définition et dans l'ensemble de mots ou expressions qu'ils subsument. C'est que les marqueurs du discours sont très labiles.

4.1.1 Caractère labile des marqueurs du discours

Discours oral vs. discours écrit Une première distinction dans l'ensemble des termes qualifiés de marqueurs du discours provient de la séparation des travaux entre discours oral et discours écrit. Chez Schiffrin (1987), *discourse* dans *discourse markers* renvoie au discours oral. De ce fait, elle inclut dans l'ensemble des marqueurs du discours les expressions verbales *you know*, *I mean* et les interjections *okay*, *oh*, *well*. Dans son modèle constitué

de cinq *plans de discussion*, un des plans concerne l'alternance des rôles entre le locuteur et l'allocutaire et la façon dont celle-ci est signalée par des marqueurs du discours tels que *well, and, but, so, or, you know*. Par exemple, *but* signale que l'allocutaire souhaite prendre le tour de parole ; *so* signale l'achèvement d'un tour de parole ; *and* signale que le locuteur poursuit son tour de parole. A cela s'ajoute le *cadre de participation* qui explique comment le locuteur et l'allocutaire s'organisent dans l'échange des tours de parole et également comment ils réfèrent au discours précédemment introduit, par exemple en utilisant le discours direct ou indirect. *oh, well, so, now, I mean, you know* sont des marques du cadre de participation. Par exemple *well* est utilisé par le locuteur pour prévenir que sa réponse ne doit pas être interprétée comme une réponse directe mais comme une digression avant que la réponse pertinente soit donnée. Toutes ces expressions jouent un rôle particulier dans la gestion des tours de parole et sont au final exclues des travaux sur le discours écrit.

Certains mots ou expressions sont, au contraire, communément admis comme des marqueurs du discours. Il s'agit des conjonctions de coordination et des conjonctions de subordination, ainsi que les locutions adverbiales ou les adverbes (Javez et Rossari, 1998), qui finalement peuvent jouer des rôles similaires tant au niveau de l'écrit que de l'oral.

Pas de caractérisation morpho-syntaxique possible Les marqueurs du discours ne peuvent pas être décrits en termes morpho-syntaxiques étant donné que ces derniers n'appartiennent pas à une seule catégorie morpho-syntaxique. Ces derniers appartiennent à plusieurs catégories morpho-syntaxiques telles qu'énumérées ci-dessus. Mais tous les éléments de ces catégories ne sont pas des marqueurs du discours.

Marqueurs vs. indices Les termes permettant de nommer les classes de mots qui nous intéressent sont construits soit avec le terme *marqueur* ou *connecteur* (*marker* ou *connective*), soit avec le terme *indice* (*cue* ou *clue*). Les termes de *marqueur* ou *connecteur* signifient que les mots ou expressions qu'ils subsument ont une fonction déterminée dans l'interprétation. Le terme d'*indice* est plus vague, couvrant un ensemble plus large d'éléments participant globalement à l'interprétation.

Polysémie des marqueurs du discours Ces expressions qui peuvent être utilisées comme des marqueurs du discours ont généralement d'autres emplois dans lesquels ils ne sont pas marqueurs du discours mais dans lesquels ils contribuent au contenu propositionnel des énoncés dans lesquels ils apparaissent (Hansen, 1998). C'est le cas, par exemple, de l'adverbe temporel *maintenant* représenté en (1) :

- (1) Nous vivons **maintenant** dans l'aile nord de la maison, dans les seules pièces épargnées par le cyclone (Le Clézio, *Le chercheur d'or*). (De Mulder, 2006, p. 22)
C'est maintenant que nous vivons dans l'aile nord de la maison.

Dans son emploi non-temporel (2), l'adverbe ne contribue plus au contenu propositionnel de l'énoncé mais sert à signaler que les présupposés et arguments validés dans l'énoncé antérieur à l'emploi de *maintenant* sont remis en question par l'énoncé que l'adverbe introduit (De Mulder, 2006) :

- (2) Julie et Marcel se voient souvent ces temps-ci. **Maintenant**, ça ne veut pas dire qu'ils sortent ensemble (De Mulder, 2006, p. 30).

**C'est maintenant que ça ne veut pas dire qu'ils sortent ensemble.*

Mais Hansen (1998) signale qu'il n'est pas toujours évident de distinguer l'emploi temporel de l'emploi non temporel de l'adverbe, comme illustré avec *maintenant* dans l'exemple (3).

- (3) Voilà ce que je te conseille. Maintenant, tu fais ce que tu veux. (Hansen, 1998, p. 242)

De plus, l'emploi temporel de *maintenant* en (1) n'exclut pas que celui-ci participe à la construction du modèle interprétatif, par exemple en opposant une situation présente à une situation passée.

Marqueurs du discours non suffisants Hansen (1998) parle d'un processus de grammaticalisation de l'emploi participant au contenu propositionnel à celui de marqueur du discours. Elle remarque également que plus ces derniers tendent à se grammaticaliser, plus ils semblent capables d'assumer des fonctions diverses. De ce fait, le marqueur seul (sans la prise en compte du contexte) n'est plus capable de marquer une seule relation de discours.

Hansen (1998) présente deux approches opposées, l'approche maximaliste et l'approche minimaliste des marqueurs du discours, que nous allons illustrer avec les exemples suivants du marqueur *et* (Hansen, 1998, p. 239) :

- (4) Londres est la capitale d'Angleterre, **et** Paris est la capitale de la France.
 (5) Hélène s'est mariée, **et** elle a eu un enfant.
 (6) Stéphane a passé la soirée à draguer toutes les filles, **et** Chantal ne veut plus lui parler.
 (7) Refais ça, **et** je te casserai la gueule.
 (8) Christophe est socialiste, **et** il a voté pour Chirac en 1995.

D'un côté l'approche maximaliste a pour objectif de décrire toutes les variations de sens d'un item lexical, reconnaissant dans les exemples ci-dessus le *et* logique dans (4), une narration dans (5), un résultat dans (6), une condition dans (7) et enfin une concession dans (8). De l'autre côté, l'approche minimaliste tente d'isoler un noyau commun à tous les emplois du marqueur du discours, comme Grice (1975) qui analyse *et* comme une implicature conversationnelle ou comme Gómez-Txurruka (2003) qui analyse le marqueur du discours *and* comme un marqueur de relation coordonnante.

Ces deux approches sont critiquées par (Hansen, 1998). D'un côté elle reproche à l'approche maximaliste de confondre le sens propre de l'expression considérée avec celui des contextes dans lesquels elle apparaît. Au niveau de l'interprétation, il n'est pas toujours clairement expliqué comment le lecteur ou allocutaire fait la différence entre les différents homonymes d'un même marqueur du discours. Il apparaît que l'interprétation du marqueur du discours va être obtenue à partir des autres mots, du contenu des énoncés reliés. De plus, elle signale également qu'analyser les différentes occurrences d'un marqueur du discours comme homonyme ne permet pas de rendre compte des glissements de sens.

De l'autre côté elle reproche aux approches minimalistes de fournir des descriptions trop générales et abstraites. De plus ces approches ne permettent pas, selon elle, de rendre compte des aspects diachroniques, i.e. du fait que les items lexicaux évoluent dans le temps, que certains sens sont ajoutés tandis que d'autres deviennent obsolètes. Ces changements de sens se trouvent absorbés dans le sens plus global du marqueur de discours considéré. Allant dans le même sens, elle note que les études sur l'acquisition du langage ont montré que l'enfant acquiert certains sens d'un item lexical avant d'autres, montrant ainsi que tous les usages potentiels d'un item ne sont pas au même niveau.

Elle propose une troisième approche, l'approche polysémique, qui cherche à pallier ces différents problèmes. Selon cette approche, on considère effectivement que les mots ont plusieurs sens qui ne dépendent pas uniquement de la pragmatique :

but rather than being homonymous and discrete, these various senses are related in an often non-predictable, but nevertheless motivated way, either in a chain-like fashion through family resemblances, or as extensions from a prototype (Hansen, 1998, p. 240-241).

L'objectif est de maintenir l'existence d'un noyau de sens tout en conservant une précision relative de la description des sens dérivés. Cet objectif est obtenu en tenant compte de l'interaction entre le sens commun d'une expression et le contexte dans lequel celle-ci apparaît, plutôt que de décrire ces sens dérivés comme des traits codés de l'expression elle-même. Toutefois, il faut noter que parfois un sens dérivé devient une partie du sens codé de l'expression.

Pour décrire les emplois des connecteurs, Rossari (2000) revendique une approche sémantique. Par opposition aux types d'analyse présentés ci-dessus qui cherchent à caractériser les contributions sémantiques pouvant être associées aux connecteurs et marqueurs du discours, Rossari (2000) cherche à dégager les contraintes sémantiques stables que les connecteurs exercent sur les contextes linguistiques gauche et droite et à mettre au jour la dissimilation du sémantique et du contextuel dans la détermination du sens de ces connecteurs.

Marqueurs du discours non nécessaires Dans la majorité des études, il est admis que les marqueurs du discours signalent comment un énoncé doit être interprété en lien avec un autre énoncé (Knott, 1996; Moore et Pollack, 1992). Il est aussi noté qu'une relation entre deux énoncés peut être interprétée sans qu'il y ait présence d'un marqueur du discours (Fraser et Malamid-Makowski, 1996; Knott, 1996; Lascarides et Asher, 1993; Rossari, 2000) et qu'ils peuvent dans certains cas être supprimés sans affecter la grammaticalité de la phrase hôte (Hansen, 1998; Knott, 1996; Rossari, 2000), comme illustré en (9) et (10).

(9) Anna s'endormit. **Puis** elle se mit à ronfler. (Le Draoulec et Bras, 2006, p. 225)

(10) Anna s'endormit. Elle se mit à ronfler. (Le Draoulec et Bras, 2006, p. 224)

La relation de discours à l'œuvre, *Narration* est mise en évidence en l'absence de prise en compte du marqueur *puis*. Les auteurs signalent qu'il est difficile de mettre en évidence le rôle du marqueur dans les cas où il est possible d'inférer la relation de discours en son absence.

Mais Pacelli-Pekba (2003) souligne à juste titre que tous les marqueurs du discours ne sont pas subsidiaires et facultatifs, comme par exemple le marqueur du discours *de toute façon* (Rossari, 2000, p. 83) :

(11) Max a oublié de se rendre à la réunion.

$$\left\{ \begin{array}{l} \text{De toute façon le comité a décidé d'ajourner cette réunion.} \\ \text{Le comité a décidé d'ajourner cette réunion.} \end{array} \right.$$

La relation de discours, nommée rétro-interprétation (Rossari, 2000), n'est ici uniquement accessible que par le biais du connecteur *de toute façon*.

Certains marqueurs du discours, comme le soulignent Wilson et Sperber (1990) et plus récemment Reboul et Moeschler (1998) imposent «de tirer de la connexion discursive des conclusions qui ne seraient pas tirées en leur absence» (Reboul et Moeschler, 1998, p. 77) comme nous le voyons avec l'exemple (12).

(12) Bill écrivit un roman. John construisit un bateau à voile.

Cet exemple peut donner lieu à deux interprétations disjointes : au niveau temporel, soit on interprète une relation de séquentialité stricte : les événements se sont déroulés dans l'ordre de la lecture ; soit on interprète une relation de simultanéité. Il n'y a pas d'indices qui permettent de considérer une interprétation plus pertinente que l'autre.

Selon Pacelli-Pekba (2003), le connecteur a dans ce cas un rôle *nécessaire de pertinence*, c'est-à-dire que celui-ci permet dans ce cas précis d'exercer une contrainte sur l'explicitation de l'une ou l'autre des deux interprétations suivant l'intention communicative du locuteur, *avant que* ou *après que* pour la séquentialité et *pendant que* pour la simultanéité :

(13) Bill écrivit un roman **avant que** John construisit un bateau à voile.
 John construisit un bateau à voile **après que** Bill écrivit un roman.
 Bill écrivit un roman **pendant que** John construisit un bateau à voile.

Dans la grande variété qui caractérise les marqueurs du discours, nous voyons que ceux-ci ne sont, le plus souvent, ni nécessaires, ni suffisants. Nous allons, dans les sous-sections suivantes, voir la place et le rôle qui leur ont été attribués dans les théories du discours (parmi lesquelles celles que nous avons présentées dans le chapitre 1).

4.1.2 Rôle des marqueurs du discours dans les théories du discours

Dans cette section, nous abordons la question du rôle et des fonctions attribués aux marqueurs du discours dans les théories. Nous allons voir que la prise en compte de ces marqueurs varie considérablement d'une théorie à l'autre, certainement une conséquence de leur caractère labile.

Leur rôle le plus décrit est celui des fonctions discursives qu'ils occupent dans la mise en œuvre des relations de discours. C'est cette fonction des marqueurs du discours qui nous préoccupe dans cette thèse. Nous nous appuyons sur les travaux de Pacelli-Pekba (2003) qui distinguent les approches dites *filtres-bouchons* et les approches dites *clé-moteur*. Dans les approches filtres-bouchons, les marqueurs du discours occupent une place secondaire. Il

est considéré que ceux-ci ne sont pas indispensables pour la détermination des relations de discours et que les relations de discours sont inférables indépendamment de toutes formes linguistiques. Dans les approches clé-moteur, les marqueurs du discours occupent une place centrale et sont considérés comme des facteurs nécessaires et suffisants pour déterminer les relations de discours.

Position zéro de la RST

Les deux caractéristiques des marqueurs du discours, ni suffisants, ni nécessaires, ont amené Mann et Thompson (1986) à les écarter de leur théorie en prônant une approche plus englobante de la structure rhétorique.

Dans la RST ne sont pas explicitée(s) la ou les façon(s) dont le lecteur reconnaît les intentions du locuteur et les effets que celui-ci souhaite produire. Les auteurs de cette théorie écartent volontairement la prise en compte dans le modèle des marqueurs du discours sans nier leur existence et leur rôle éventuel dans l'interprétation.

Marqueurs de segmentation

Pour accéder à la structure d'un discours, la première étape consiste à découper le discours en segments. Les marqueurs du discours constituent une source d'information signalant des frontières de segments. À l'oral, nous avons vu que les marqueurs du discours signalent l'alternance des tours de parole. Schiffrin (1987) est la première à notre connaissance à faire un inventaire du rôle des marqueurs du discours pour la tâche de segmentation. Les marqueurs du discours ne participent évidemment pas uniquement à cette tâche et jouent un rôle à tous les plans de discussion décrits dans ce cadre.

À l'écrit, certains marqueurs du discours ont été décrits du point de vue de leur rôle dans la délimitation des segments. Le rôle de ces marqueurs dans la théorie de Grosz et Sidner (1986) intervient également au niveau de la *structure linguistique* qui concerne les séquences d'énoncés.

Dans l'exemple illustré figure 4.1, les énoncés sont représentés de (a) à (p). Des groupes d'énoncés forment les segments de discours DS (*Discourse Segment*). L'utilisation des marqueurs du discours pour indiquer les frontières de ces DS est illustrée en (i) et (k). Dans l'énoncé (i), le marqueur *in the first place* signale le début du segment de discours DS5 tandis que le marqueur *moreover* signale la fin du segment DS5 et le début du segment DS6. Ces marqueurs du discours apportent également des informations au niveau de la structure intentionnelle, à savoir que les segments DS5 et DS6 sont dominés par le segment DS4.

Une étude plus récente et appliquée (Piérard et Bestgen, 2006) propose une technique automatique pour l'identification des expressions linguistiques qui signalent la structure des textes, plus particulièrement la présence de discontinuité thématique. Ils partent ainsi de l'hypothèse que la structure du discours est signalée par des éléments linguistiques surfaciques et entreprennent ce qu'ils appellent une «pêche» aux marqueurs linguistiques de la discontinuité thématique en s'appuyant sur le découpage en paragraphe et la cohésion lexicale (calculée automatiquement). Dans cette approche, les marqueurs du discours (toutes

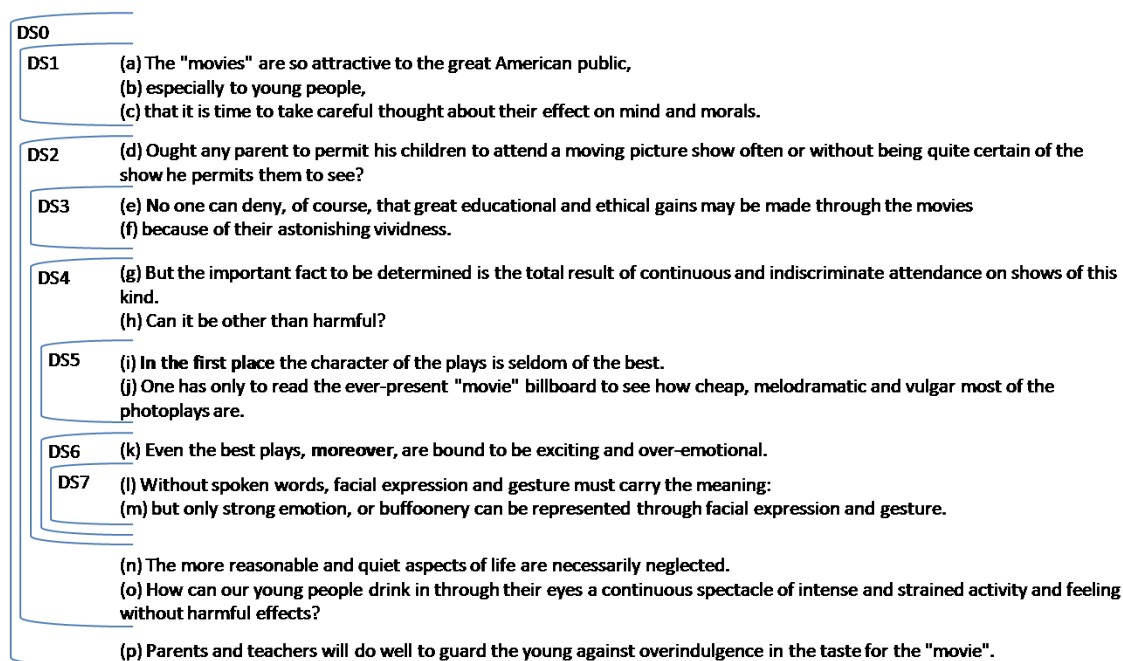


FIG. 4.1 – Un exemple 'Movies Essay' (Grosz et Sidner, 1986)

les expressions linguistiques signalant la structure des textes), de la même façon que les changements de paragraphe, sont considérés par les auteurs comme des moyens employés par les scripteurs pour signaler la présence d'une discontinuité thématique.

Dans cette section, nous avons présenté des théories et études qui ont considéré les marqueurs du discours du point de vue de la délimitation de segments de discours. Le rôle des marqueurs du discours n'est évidemment pas limité à cette tâche et ceux-ci sont impliqués à d'autres niveaux de la cohérence discursive. Par exemple, ils signalent la façon dont ces segments de discours sont connectés les uns avec les autres. Les études présentant les corrélations existant entre marqueurs du discours et relations de discours sont nombreuses.

Nous présenterons dans la section suivante des théories dans lesquelles les marqueurs du discours sont considérés comme secondaires. Leur rôle est de «filtrer» ou «bloquer» en quelque sorte des interprétations possibles. Nous présenterons dans les prochaines sections des travaux spécifiques sur la corrélation entre indices de surface et relations de discours (section 4.2.2) et sur des marqueurs du discours de la relation d'*Élaboration* (section 4.4.2), qu'ils s'inscrivent ou non dans un cadre théorique.

Marqueurs, «filtres-bouchons» des relations de discours

Contrairement au modèle de la RST sur lequel ils s'appuient, Maier et Hovy (1993) prennent en compte le rôle des marqueurs de discours dans la cohérence discursive. Le marqueur du discours occupe une place importante dans leur théorie sans toutefois être

à la base du modèle. Selon Maier et Hovy (1993), les marqueurs du discours renforcent les relations de discours comme dans l'exemple suivant où les segments (a) et (b) sont reliés par la relation d'*Élaboration*. En effet, le lecteur peut reconnaître que l'information apportée en (b) spécifie en ajoutant des détails l'information donnée en (a). La relation est de plus signalée par le marqueur du discours «for example» :

- (14) (a) A higher-level language can be used on more complex symbol structures, such as lists, trees and character strings. (b) In LISP, **for example**, the contents of a number of storage cells in the underlying abstract machine can be interpreted together as representing a list of items. (Maier et Hovy, 1993, p. 75)

Pour le français, Rossari (2000) montre également que *par exemple* est un spécificateur de la relation d'*Exemplification* (i.e. *par exemple* exprime une relation entre deux segments de discours qui existe par ailleurs sans la présence du marqueur).

Hobbs (1985) n'inclut pas les marqueurs du discours dans les définitions des relations de discours mais il leur accorde un rôle majeur dans la reconnaissance des relations en considérant que leur insertion, couplée avec un test de commutation des segments, peut confirmer l'inférence d'une relation de discours :

If we can insert *then* between S_0 and S_1 , and the sense would be changed if we reversed the segments, then the occasion relation is an excellent candidate.
If we can insert *because*, the explanation relation becomes a strong possibility.
(Hobbs, 1985, p. 32)

Les marqueurs du discours sont utilisés comme des tests pour reconnaître les relations de discours. Cependant ce type de test ne peut pas à lui seul engager l'inférence d'une relation de discours car l'insertion d'un marqueur du discours peut modifier la relation de discours, comme le montrent ces exemples de Wilson et Sperber (1990) :

- (15) Pierre n'est pas stupide. Il sait rentrer tout seul chez lui.
a) Pierre n'est pas stupide. *Après tout*, il sait rentrer tout seul chez lui.
b) Pierre n'est pas stupide. Il sait *donc* rentrer tout seul chez lui.

L'énoncé en (15) a deux interprétations possibles disjointes : dans la première interprétation, la deuxième phrase apporte un argument pour confirmer l'assertion de la première. Cette relation est confirmée par l'insertion du marqueur «après tout» en (a). Dans la deuxième interprétation, l'insertion du marqueur «donc» en (b) explicite l'interprétation de l'état de la deuxième phrase comme étant un résultat de celui de la première.

Dans le cadre de la théorie de la Pertinence, Wilson et Sperber (1990) définissent les connecteurs comme des «filtres» encodant des contraintes d'interprétation. De plus, ils considèrent les marqueurs du discours comme permettant de réduire l'effort de traitement menant à la compréhension des discours. Comme nous l'avons vu, l'exemple (15) a deux interprétations : soit l'interprétation (a) d'*Evidence* mise en exergue par le marqueur «après tout» ; soit l'interprétation (b) de *Résultat* mise en évidence par le marqueur «donc». Dans ce cas, le marqueur permet de réduire l'effort de traitement du destinataire en indiquant le type de processus inférentiel à mettre en place.

Dans le cadre de la SDRT, les relations de discours sont de nature sémantique. Comme nous l'avons vu précédemment, les marqueurs du discours ne sont pas systématiquement et exhaustivement pris en compte dans le modèle (Asher, 1993; Asher et Lascarides, 2003) mais il existe, au sein de ce cadre théorique, des règles d'inférence qui s'appuient sur l'emploi, entre deux propositions, d'un marqueur de discours tel que *and then* pour la relation de *Narration* (Asher et Lascarides, 2003, p. 202). Concernant les relations non marquées (lorsque deux énoncés sont reliés par une relation de discours sans la présence d'un marqueur du discours), d'autres règles d'inférence basées sur des indices de sémantique lexicale des éventualités décrites par les énoncés et des connaissances du monde sont convoquées. Il reste à examiner la façon dont ces deux types d'information interagissent dans la mise en œuvre des relations de discours.

De récentes contributions ont également étudié le rôle de certains marqueurs lexicaux dans la mise en œuvre des relations de discours et ont proposé des règles «filtres» à partir de ceux-ci pour l'inférence ou le blocage de relations de discours. Nous pouvons citer à titre d'exemple les travaux sur *alors* qui autorise *Narration* et déclenche une relation de *Résultat faible* (Le Draoulec et Bras, 2007; Bras *et al.*, 2009), *d'abord* (Bras, 2007; Bras et Le Draoulec, 2009)¹, ou sur *puis* qui déclenche la relation de *Narration* et bloque celle de *Résultat* (Bras *et al.*, 2001).

Bras et Le Draoulec répondent aux remarques de Rossari (2000) (selon laquelle les approches conceptuelles du type SDRT ne rendent pas justice au rôle des marqueurs du discours car le sémantisme de ces derniers est absorbé par les relations de discours) en montrant que leurs analyses de *puis* et *alors* (entre autres) amènent à considérer des versions faibles et fortes des relations de discours (Bras *et al.*, 2009). Mais l'ajout de nouvelles relations pour la description de l'emploi de chaque connecteur ne les satisfait pas et elles souhaiteraient ré-analyser les connecteurs temporels en conservant un modèle de relations de discours définies minimalement et en formalisant les contraintes apportées par chaque connecteur comme des paramètres de la relation².

Par opposition aux approches «filtres-bouchons», les approches dites «clé-moteur» considèrent les marqueurs du discours comme la *clé* ou le *moteur* pour accéder aux relations de discours. Martin (1992) et Knott (1996) adoptent ce type d'approche.

Marqueurs du discours, «clé-moteur» pour établir un ensemble de relations de discours

Martin (1992) se base sur les instructions données par les connecteurs pour déterminer un ensemble de *relations conjonctives*. Il distingue quatre familles de relations : les relations additives, les relations comparatives, les relations temporelles et les relations consécutives. La classification des relations qu'il obtient est basée sur les caractéristiques des connecteurs : il distingue ainsi les relations paratactiques des relations hypotactiques et les relations internes (qui s'ancrent sur les liens entre les actes illocutoires véhiculés par les énoncés d'un texte) des relations externes (qui s'ancrent sur des liens entre les éventualités décrites dans ces énoncés). De plus, il considère que s'il existe une connexion implicite entre deux

¹Nous reviendrons sur ce travail qui porte, entre autres, sur la relation d'*Élaboration*.

²Communication personnelle.

segments, celle-ci sera mise en évidence par la possibilité d'introduire un connecteur entre ces deux segments. Les travaux de Martin se concentrent sur la taxonomie des connecteurs tandis que Knott (1996) propose d'utiliser les marqueurs du discours pour motiver un ensemble de relations de discours.

Knott (1996) part du postulat suivant :

Evidence that a relation is actually used by speakers of a language can be obtained by looking at the language itself. (Knott, 1996, p. 3)

L'argument principal de l'auteur est la corrélation entre la présence d'un marqueur du discours et l'existence d'une relation de discours, car, selon lui, les relations modélisent les constructions psychologiques en représentant le processus mental à l'oeuvre lorsque les humains produisent et interprètent des textes. C'est parce que les humains produisent des textes que le langage contient les ressources pour rendre les relations explicites.

Knott (1996) présente une méthodologie qui se base sur des mots-clés (*cue phrases*) pour la validation d'un ensemble de relations de discours. Cette méthode a pour avantage de considérer dès le départ des données linguistiques concrètes. Les marqueurs du discours deviennent donc une source commode pour valider un ensemble de relations. L'ensemble des mots-clés pouvant jouer un rôle dans la structuration du discours est large et non homogène. Knott (1996) propose donc une méthodologie du repérage des mots-clés opérant des liens dans le discours, appelés *relational phrases*, basée sur l'application de plusieurs tests :

1. Isoler l'expression à tester (en gras) et sa proposition d'accueil (non barrée) :
~~*John and Bill were squabbling. John fell*~~ ***because*** *Bill pushed him* ~~*∴ that was how it all started*~~
2. Substituer les éléments anaphoriques et cataphoriques avec leur antécédent («him» devient John) et inclure les éléments élidés :
because *Bill pushed John*
3. Le texte obtenu après ces deux manipulations paraît incomplet (il manque un ou plusieurs segments supplémentaires pour assurer la cohérence de l'ensemble), donc l'expression que nous testons est peut-être un marqueur du discours.
4. Enfin, si après suppression du mot-clé que nous testons, la proposition obtenue n'apparaît plus comme incomplète, nous pouvons conclure que notre mot-clé candidat est un marqueur du discours :
Bill pushed John

Ainsi, les auteurs obtiennent une liste de marqueurs du discours qui se divisent en quatre classes syntaxiques :

- Les coordinateurs : *but, and...*
- Les subordonateurs *before, after...*
- Les adverbes conjonctifs *As a consequence, however...*
- Les expressions phrastiques *It follows that, it may seem that...*

Enfin, Knott propose de distinguer quatre types de relations entre les marqueurs : les marqueurs synonymes qui peuvent apparaître dans le même contexte ; les marqueurs exclusifs qui ne peuvent jamais apparaître dans le même contexte ; les marqueurs hypéronymes ou

hyponymes : un marqueur est un hypéronyme d'un autre marqueur lorsqu'il est possible de remplacer le marqueur hyponyme par le marqueur hypéronyme et que l'inverse n'est pas forcément possible.

Knott (1996) s'appuie sur les travaux en psycholinguistique de Sanders *et al.* (1992) qui mettent en avant l'importance des aspects psychologiques dans la construction d'un modèle de la cohérence. Sanders *et al.* (1992) et Knott (1996) perçoivent les relations comme des constructions psychologiques, lesquelles permettent au lecteur de construire une représentation cognitive du discours. Cependant les travaux de Sanders *et al.* (1992) diffèrent de ceux de Knott (1996) en ce qui concerne l'importance qui est accordée aux marqueurs du discours. Bien que Sanders *et al.* (1992) reconnaissent que les marqueurs du discours jouent un rôle important dans la mise en oeuvre de la cohérence discursive, ils ne proposent pas qu'un unique marqueur du discours existe pour chaque relation.

Les approches psycho-linguistiques de la cohérence discursive se distinguent des approches précédemment présentées dans le sens où ces travaux ont pour objectif d'expliquer plus spécifiquement pourquoi les marqueurs du discours apparaissent. Les marqueurs du discours rendent explicites les relations de discours. D'un point de vue psychologique, il est montré que sans la présence d'un item lexical explicite, le lecteur présentera des difficultés à établir une représentation cognitive du discours. Pour le cas des connecteurs de causalité, des études ont montré d'une part que le processus de traitement des discours se fait plus rapidement en présence des connecteurs (Sanders et Noordman, 2000) et d'autre part que la présence de connecteurs de causalité dans un texte améliore significativement la compréhension de celui-ci (Degand et Sanders, 2002).

Dans l'approche suivante, les marqueurs du discours jouent également un rôle «clé-moteur» selon l'idée que l'accès à la structure du discours peut se faire à travers les marqueurs du discours.

Les marqueurs du discours dans les approches «clé-moteur» sont considérés comme nécessaires et suffisants. Ils sont le «moteur» ou la «clé» d'accès aux relations de discours. Les auteurs développant ce type d'approche postulent une systématique mécanique entre le sens des connecteurs et celui des relations de discours, mais cette systématique n'apparaît pas évidente Pacelli-Pekba (2003).

D'abord elle souligne qu'il existe des relations de discours qui n'ont pas de connecteurs spécifiques soulevant la question de la légitimité de baser une typologie des relations de discours sur la base d'une taxinomie des connecteurs (voir pour illustration le cas de la relation d'*Elaboration Object-Attribute* de la RST et *Élaboration d'entité* de la SDRT). Ensuite elle note que dans l'activité de communication, le locuteur ne choisit pas toujours les connecteurs explicites d'une relation de discours (i.e. ceux dont le sens est directement corrélé avec le concept d'une et une seule relation de discours) mais au contraire il choisit généralement les connecteurs les plus généraux (i.e. ceux dénotant une pluralité de sens pour exprimer les relations de discours). Parmi les connecteurs discursifs relevés dans le projet LEXCONN, 73 connecteurs (soit 23,7%) sont compatibles avec plusieurs relations de discours (Roze *et al.*, 2010), confirmant l'hypothèse qu'un grand nombre de connecteurs sont ambigus.

Marqueurs, base d’annotation L’approche que nous allons présenter maintenant est entièrement basée sur des corpus et donne une place centrale aux marqueurs du discours (appelés *connectives*). Le projet du *Penn Discourse Tree Bank* (PDTB) est un projet d’annotation manuelle au niveau discursif (Prasad *et al.*, 2006, 2007, entre autres). Leur objectif est de fournir un corpus annoté des marqueurs du discours et des segments (appelés arguments) qu’ils relient.

Les auteurs distinguent pour ce travail trois types de marqueurs du discours qu’ils nomment *explicit connectives* : les conjonctions de subordination, les conjonctions de coordination et les adverbiaux connecteurs. Une originalité de cette étude est de considérer, en plus des connecteurs explicites, ce qu’ils appellent des connecteurs implicites qui sont annotés lorsqu’il est possible d’intégrer un marqueur du discours entre deux segments.

Lors de la tâche d’annotation manuelle, les annotateurs doivent annoter en premier lieu les marqueurs du discours, puis annoter les arguments que ces marqueurs du discours permettent de relier. En ce qui concerne les connecteurs implicites, ceux-ci sont annotés par l’annotateur une fois que celui-ci a pris connaissance du texte qu’il annote et dans la mesure où il est capable d’intégrer un connecteur.

Enfin, les annotateurs doivent spécifier la relation de discours qui est vérifiée entre les deux arguments en s’appuyant sur la sémantique du connecteur.

Pour conclure, il nous paraît important de noter que la majorité des études s’accordent sur le fait que les marqueurs du discours facilitent la reconnaissance qu’une relation de discours existe entre deux segments et qu’ils constituent un point d’accès à la structure discursive. Nous proposons, dans la section suivante, de définir les notions d’indices, de marqueurs de discours et de marqueurs de relation.

4.1.3 Synthèse sur indices et marqueurs du discours

Les marqueurs du discours (ou *marqueurs*) donnent l’indication au lecteur qu’une relation de discours existe entre deux segments. Le langage contient d’autres ressources pour faciliter l’interprétation des textes que nous appelons *indices*, ensemble très large d’éléments participant à l’interprétation. Les marqueurs du discours, ensemble plus restreint d’éléments pour lesquels la ou les fonctions discursives ont été établies, peuvent être, dans certains cas, des marqueurs d’une (et une seule) relation de discours. Mais généralement un marqueur d’une relation de discours fait intervenir un faisceau d’indices.

Lors de l’analyse des indices, comme marqueurs potentiels de la relation d’*Élaboration* (Partie II de la thèse), nous analyserons leur fonction discursive, avec pour objectif de déterminer, d’une part, si ceux-ci contribuent :

1. à produire³ la relation de discours (i.e. à mettre en place une relation entre deux segments de discours qui n’existe pas sans la présence du marqueur) : c’est le cas par exemple avec le marqueur du discours *de toute façon*, Cf. exemple (11) ;
2. à spécifier⁴ la relation de discours (i.e. à exprimer une relation de discours entre deux segments de discours qui existe par ailleurs sans la présence du marqueur), comme dans l’exemple ci-dessous :

³Rossari (2000) parle de connecteur producteur.

⁴Rossari (2000) parle de connecteur spécificateur.

(16) Max était fatigué donc il est rentré tôt. (Rossari, 2000, p. 27)

3. à bloquer l'inférence de la relation de discours, comme l'ont montré (Bras *et al.*, 2001) avec le connecteur *puis* qui bloque l'inférence de la relation de *Résultat*. L'interprétation de *Résultat* est possible dans (a) mais plus dans (b) à cause de la présence du connecteur *puis* :

- (17) a. L'acide tomba dans le liquide. Une explosion se produisit.
b. L'acide tomba dans le liquide. **Puis** une explosion se produisit.

D'autre part, nous déterminerons les contextes particuliers (traduits au moyen d'un faisceau d'indices) dans lesquels nos marqueurs potentiels sont considérés comme des marqueurs de la relation d'*Élaboration*.

4.2 Autres indices de la signalisation discursive

4.2.1 Des indices moins étudiés

D'autres indices disponibles à la surface des textes jouent un rôle dans la signalisation de l'organisation discursive, tels que les structures syntaxiques (Delin *et al.*, 1996), (Scott et De Souza, 1990) et (Grote *et al.*, 1995), mais ceux-ci ont été bien moins étudiés que les marqueurs du discours.

Le travail de Reichman (1978) a consisté à dégager des régularités accompagnant les transitions entre *context spaces*. Il a montré que ces transitions se caractérisent par l'emploi de manifestations linguistiques particulières, dont les mots-clés (*clue words*) mais également des expressions déictiques, des expressions introduisant un topique (*What I was saying before*, *Speaking of X*), des procédés de référence et des changements de temps verbaux.

Harabagiu (1998); Harabagiu et Maiorano (1999) et Harabagiu (1999) tiennent compte, quant à eux, d'une autre source d'information, la cohésion lexicale, pour reconnaître les relations de discours. Ils partent de l'hypothèse que la cohésion lexicale est un indice de surface de la cohérence du texte et que celle-ci indique les connaissances lexico-sémantiques sur lesquelles se basent l'inférence des relations de cohérence. A partir du texte, une chaîne cohésive avec des informations extraites de WordNet et une chaîne de coréférence sont construites sur lesquelles se basent ensuite les inférences des relations de discours. Nous reviendrons sur cette méthodologie de détection automatique des relations de discours dans le chapitre 10 dans lequel il sera question d'utiliser des indices de cohésion lexicale pour l'identification automatique de la relation d'*Élaboration*.

4.2.2 Corrélation entre indices et relations de discours

Les travaux qui étudient la corrélation entre indices et relations de discours sont nombreux. Nous pouvons distinguer deux types d'étude, d'une part les études qui partent d'un indice pour déterminer la ou les relation(s) que cet indice permet de signaler (approche sémasiologique que nous illustrerons dans les chapitres 5, 6 et 7) et d'autre part les études qui partent d'une relation de discours afin d'appréhender ses réalisations linguistiques dans les textes (approche onomasiologique que nous illustrerons dans le chapitre 9).

Les travaux proposés par Delin *et al.* (1996) concernent ce deuxième type d'étude. Ils proposent une étude sur les correspondances entre contenu sémantique et réalisation syntaxique et lexicale dans un corpus de textes instructionnels en anglais, français et portugais dans une perspective de génération automatique de texte. Ils présentent un éventail des réalisations syntaxiques et lexicales de deux relations, *Generation* et *Enablement* en prenant pour appui les définitions de Goldman (1970).

Les auteurs disposent d'un corpus de 65 exemples pour chacune des relations et pour chaque langue étudiée. Leur objectif est d'observer comment les arguments de ces relations sont exprimés dans les textes (constructions syntaxiques et marqueurs du discours) et les contraintes qui pèsent sur l'ordre des arguments, pour montrer enfin que les choix disponibles sont contraints sémantiquement, syntaxiquement et pragmatiquement. Pour chaque argument de la relation, les auteurs relèvent le type syntaxique de chaque proposition (impérative, infinitive, gérondive...) et étudient les contraintes qui pèsent sur l'ordre de ces propositions (par exemple, en français, pour la relation de *Generation*, les propositions gérondives exprimant le but n'apparaissent qu'en seconde position pour l'expression du segment *générant*). Ils relèvent que l'expression des relations de discours est fortement corrélée au choix d'une structure syntaxique dans les trois langues.

Concernant les marqueurs du discours, les auteurs notent également une forte corrélation entre l'expression d'une relation de discours et l'apparition d'un marqueur du discours. Ils notent également de larges différences entre les langues étudiées et les occurrences des marqueurs du discours par relation. En portugais, toutes les relations de discours doivent être explicitement marquées tandis que la moitié des relations de discours sont marquées en français et 37 des 130 propositions le sont en anglais.

Ce travail nous concerne particulièrement car il propose une étude qui part de relations de discours pour décrire comment ces relations sont réalisées dans des corpus. Nous proposerons une analyse similaire dans le chapitre 9 pour la relation d'*Élaboration* en considérant des exemples de la relation extraits du corpus annoté ANNODIS pour voir les réalisations syntaxiques et lexicales de ladite relation. Ce travail se distingue cependant de notre approche par le type de texte et par l'approche multilingue.

Identifier des réalisations linguistiques d'une relation de discours est rendu nécessaire pour accéder à celle-ci de façon automatique. Lorsque les théories de la cohérence discursive sont invoquées dans des applications de TAL, une place importante est donnée au rôle des indices surfaciques de la signalisation discursive, étant donné que ceux-ci sont plus facilement repérables automatiquement. Le repérage systématique des relations de discours requiert un inventaire aussi riche que possible des indices disponibles pour le repérage automatique.

4.3 Rôle des indices de la signalisation discursive en TAL

Nous présentons, dans les deux sous-sections suivantes, le rôle que vont jouer les marqueurs du discours dans deux applications de TAL, la génération automatique de texte et le résumé automatique.

4.3.1 La génération automatique de texte

Les études dans le domaine de la génération automatique de texte sont fortement influencées par les approches psycholinguistiques (Scott et De Souza, 1990) pour ce qui concerne l'utilisation des marqueurs du discours. Nous avons déjà présenté l'intérêt porté par les études en génération de texte aux théories de la cohérence discursive. Cet intérêt est lié aux besoins de la discipline qui, pour produire des textes cohérents, doit rendre explicite la structure rhétorique sous-jacente, en expliquant, prédisant et générant relations de discours et marqueurs du discours. Les relations de discours sont employées pour planifier des textes ou des réponses (systèmes de question-réponse) cohérents. Le but des approches de génération de texte «is to describe a method of planning paragraphs to be coherent while avoiding unintended spurious effects that result from the juxtaposition of unrelated pieces of text» (Hovy, 1988a).

(Scott et De Souza, 1990) insistent sur le rôle des marqueurs du discours et des structures syntaxiques pour rendre explicites dans le texte généré les relations de discours sous-jacentes. Planifier un texte avec des relations de discours permet, entre autres, de guider des choix de surface, par exemple la sélection de connecteurs, ce qui permet d'améliorer la cohésion des textes produits (McKeown, 1985).

De nombreuses études en génération de texte se concentrent sur la sélection d'un marqueur de discours et la position que celui-ci va occuper dans le texte généré (Hovy, 1993; Di Eugenio *et al.*, 1997). Il est généralement accepté que là où se tient une relation de discours, un marqueur de discours peut, dans un grand nombre de cas, être utilisé. D'autres études ont de plus suggéré que la mise en page (Bouayad-Agha et Scott, 1998), la ponctuation (Dale, 1991) et le genre de texte (Grote *et al.*, 1995) vont avoir une influence sur les marqueurs de discours nécessaires et/ou attendus. Enfin, il est admis que, dans le domaine de la génération de texte, il est extrêmement utile d'établir des contraintes pour savoir s'il faut employer un marqueur de discours et si oui lequel.

Cependant, les travaux en génération automatique de texte ont dégagé les limites de l'association entre relations de discours et marqueurs du discours, montrant que plusieurs marqueurs signalant la même relation de discours devaient être distingués selon d'autres critères d'ordre pragmatique.

Dans une étude de l'emploi des marqueurs de discours *since* (sens concessif de l'item), *because*, *but*, *although* pour la génération de texte, Elhadad et McKeown (1990) montrent que l'utilisation des relations de discours pour générer des marqueurs du discours n'est pas suffisante (du moins pour les approches conceptuelles du type RST ou SDRT). Par exemple, il n'est pas suffisant d'associer la relation de discours *Contraste* avec le marqueur de discours *but* car il est souvent nécessaire de distinguer différents marqueurs d'une même relation de discours. Il faut, par exemple, distinguer les marqueurs *since* vs. *because* pour la relation d'*Explication* ou *but* vs. *although* pour la relation de *Contraste*. Par conséquent les auteurs utilisent un modèle de typage des propositions à relier, appelé *Interpretative Format Model*. Les marqueurs du discours (*connectives*) sont alors définis comme un ensemble de contraintes entre les traits des propositions qu'ils connectent. Cette approche vise à distinguer les quatre marqueurs du discours *but*, *although*, *since*, *because*. Toutes les propositions à relier avec un de ces marqueurs sont alors définies par leur contenu propositionnel et par quatre critères pragmatiques : l'orientation argumentative (ensemble des

conclusions supportées par la proposition) (Ducrot, 1983) ; le statut fonctionnel (relation structurelle établie avec le segment précédent) (Roulet *et al.*, 1985) ; le trait polyphonique (indications sur le fait que le locuteur attribue ses propos à lui-même ou à une autre personne) (Ducrot, 1983) et la procédure de thématisation (description des connections (par exemple anaphoriques) entre les entités des propositions).

Comme Elhadad et McKeown (1990), Grote *et al.* (1995) reconnaissent que les arbres fournis par la RST ne sont pas à eux seuls suffisants pour fournir toutes les informations nécessaires à la décision d'une structure syntaxique et/ou d'un marqueur du discours. Grote *et al.* (1995) suggèrent qu'il faut définir deux niveaux, le niveau informationnel et le niveau intentionnel (Moore et Pollack, 1992). Au niveau informationnel, les relations s'établissent entre deux événements. Par exemple pour la relation de *Cause*, un segment décrit un événement qui est présumé être la cause de l'événement décrit dans le segment précédant. Au niveau intentionnel, les relations résultent d'une volonté du locuteur de produire un effet sur l'état mental du lecteur ou allocutaire. Par exemple pour la relation *Evidence*, un segment a pour but d'augmenter la probabilité que le lecteur ou allocutaire croit le contenu du segment précédent. Pour Grote *et al.* (1995), les marqueurs du discours et les constructions syntaxiques employés seront différents selon que la relation se situe au niveau informationnel ou au niveau intentionnel, levant ainsi toute ambiguïté sur la nature de la relation. Grote *et al.* (1995) travaillent sur la relation de *Concession* et ont justement pour objectif de créer une grammaire qui puisse être utilisée pour générer un large éventail de marqueurs du discours qui signalent les différences au sein de cette relation.

4.3.2 Le résumé automatique de texte

Les travaux de Marcu (2000) visent la détection automatique de la structure rhétorique d'un texte dans le but de dégager les segments (typiquement les noyaux des relations au sens de la RST) à conserver dans le résumé obtenu automatiquement à partir de la structure rhétorique.

L'algorithme permettant de déterminer les structures de discours valides se base sur la reconnaissance de marqueurs du discours, de liens cohésifs et sur une formalisation des contraintes pesant sur les structures de discours valides.

Marcu (2000) souligne l'importance des marqueurs du discours (qu'il appelle *connectives*) dans l'identification d'une part des segments de discours (typiquement la proposition) et d'autre part des relations de discours en s'appuyant sur l'hypothèse de Halliday et Hasan (1976) selon laquelle les marqueurs du discours sont utilisés comme des liens cohésifs entre des propositions ou des phrases adjacentes. Les mêmes marqueurs du discours peuvent signaler des relations entre des unités textuelles larges (macro-structure). Par exemple, Kintsch (1977) montre que les marqueurs du discours *and*, *so* and *but* signale des frontières entre des parties d'une histoire. Il souligne également l'utilisation massive des marqueurs du discours dans son corpus d'articles scientifiques et d'articles de journaux (approximativement un toutes les deux propositions pour l'anglais⁵) qui rend possible le repérage

⁵Nous notons que cette remarque est contraire à ce que l'on trouve dans d'autres travaux sur l'anglais : pour les textes instructionnels, Delin *et al.* (1996) relèvent 37 marqueurs pour 130 propositions ; Di Eugenio *et al.* (1997) relèvent 476 marqueurs pour 780 relations dans un corpus de tutoriels et Harabagiu (1999) relève que moins de 30% des relations de discours annotées dans un corpus de journaux (*Wall Street Journal*) sont signalées par un marqueur du discours. On ne peut conclure si ces différences de densité des

automatique de la structure discursive des textes, à condition d'employer des algorithmes de désambiguïsation de ces marqueurs du discours.

Il tient également compte de la cohésion lexicale pour l'identification des relations de discours. Par exemple, deux segments qui parlent de choses différentes seront reliés par la relation *Joint* et deux segments qui parlent de la même chose seront reliés par les relations d'*Élaboration* ou d'*Arrière-Plan*.

La dernière section de ce chapitre est consacrée à un état de l'art sur la signalisation de la relation d'*Élaboration*, dernier point à aborder avant de présenter nos travaux sur la question dans les parties 2 et 3.

4.4 La signalisation de la relation d'*Élaboration*

4.4.1 Une relation non marquée et une relation par défaut

4.4.1.1 Une relation non marquée

La relation d'*Élaboration* est souvent présentée dans les études sur les marqueurs du discours comme une relation pour laquelle il n'existe pas de marqueurs lexicaux (Knott, 1996; Scott et De Souza, 1990).

Knott (1996) constate que la relation d'*Élaboration Object Attribute* de la RST peut être vérifiée entre deux phrases alors qu'il n'est pas possible de trouver, en anglais, de marqueur du discours qui puisse spécifier la relation :

- (18) Dow Associates is one of the Britain's largest companies.
 *Indeed/*Specifically/*Furthermore/*Incidentally its head office is in Kensington,
 where Dow himself presides.

Ces contextes discursifs dans lesquels aucun marqueur du discours ne peut s'insérer sont problématiques pour l'hypothèse de Knott (1996) (qu'il existe une corrélation entre l'ensemble des marqueurs du discours et l'ensemble des relations de discours).

La solution proposée par Knott (1996) et Knott *et al.* (2001) est d'introduire d'autres mécanismes théoriques pour rendre compte des contextes dans lesquels aucun marqueur du discours n'est approprié. Nous avons déjà évoqué cette proposition dans la section 3.3.1.

Dire que la relation d'*Élaboration* est non marquée est, de notre point de vue, un postulat trop fort. Il faut d'abord préciser que le type d'*Élaboration* considéré par Knott (1996) est seulement un sous-type appelé *Object-Attribute*. Ensuite, Knott (1996) évoque uniquement les marqueurs du discours tandis que Scott et De Souza (1990), Halliday (1994), Thompson (2004) et Prévot *et al.* (2009) soulignent que certaines structures syntaxiques telles que les relatives non restrictives et les appositions sont des indices des élaborations d'entité (équivalent dans la SDRT de la relation *Object-Attribute* de la RST).

marqueurs du discours provient de la définition donnée à ces derniers dans les différents travaux ou plus simplement du genre du corpus.

4.4.1.2 Une relation par défaut

La relation d'*Élaboration* est souvent considérée comme une relation par défaut. C'est le cas dans les travaux de Marcu (2000). Comme nous l'avons déjà évoqué, ces travaux ont pour objectif le repérage automatique des relations de discours au sens de la RST. En ce qui concerne la relation d'*Élaboration*, outre le fait qu'il s'appuie sur le marqueur du discours *specifically* (peu employé dans les textes), Marcu (2000) propose une identification sur la base d'un indice non linguistique : si le nombre de phrases dans un paragraphe ou le nombre de paragraphes dans une section est petit et qu'aucun marqueur du discours n'est utilisé, alors la relation inférée entre les phrases ou entre les paragraphes est *Élaboration*. Cet indice rend typiquement compte de la considération de cette relation comme une relation par défaut.

Dans le cadre d'une étude formelle des relations de discours (Blutner et Zeevat, 2003), Jasinskaja (2010) considère la relation d'*Élaboration* comme une relation s'établissant par défaut, i.e. lorsqu'il n'y a pas d'indices linguistiques indiquant qu'une autre relation de discours est inférée. Le modèle d'inférence des relations de discours proposé par Jasinskaja (2010) s'appuie (a) sur la notion de topique de discours et (b) sur le mécanisme des implicatures de Grice (1975). Elle montre comment la combinaison de l'exhaustivité comme mode d'interprétation par défaut et le principe de continuité topicale conduit à l'inférence de la relation d'*Élaboration* (ou plus précisément à une interprétation de coréférence événementielle entre les deux éventualités en jeu), que nous illustrons avec l'exemple suivant :

- (19) (a) Alena broke her skis. (b) She lost her main transportation means. (Jasinskaja, 2010, p. 109)
Alena a cassé ses skis. Elle a perdu son moyen de transport principal. (Notre traduction)

Le principe de continuité topicale spécifie qu'en l'absence d'indications spécifiques (*coding material*) l'interprétation d'un nouveau segment se fait sous la domination du topique de discours en cours. Selon ce principe, on considère que les deux phrases de l'exemple (19) sont dominées par le même topique discursif.

L'interprétation exhaustive de (a) spécifie que «Alena breaking her skis» est le seul événement pertinent qui s'est déroulé. De même pour (b), l'interprétation exhaustive spécifie que «Alena losing her main transportation means» est le seul événement pertinent qui s'est déroulé. De ces deux constatations, (Jasinskaja, 2010) déduit qu'il s'agit donc du même événement.

Dans la présente approche, la coréférence événementielle résulte de l'application de ces deux principes par défaut. Il s'ensuit que la relation d'*Élaboration* est considérée comme une relation par défaut. Par conséquent, les autres relations telles que *Narration* ou *Contraste* ne sont pas considérées comme des relations par défaut et doivent être explicitement signalées.

La relation d'*Élaboration* est toujours inférée lorsque les deux principes d'interprétation exhaustive et de continuité topicale sont maintenus. Au moins un des deux principes doit être enfreint pour qu'une autre relation de discours soit inférée. Ces principes peuvent être annulés par des moyens linguistiques explicites qui codent (a) une relation de discours

spécifique ou (b) un changement de topique.

Jasinskaja (2010) donne deux exemples de violation des principes d'interprétation exhaustive et de continuité topicale. Le marqueur du discours *and* annule le principe de continuité topicale, excluant ainsi la lecture coréférentielle et une intonation montante indique que le discours est incomplet, annulant ainsi le principe d'interprétation exhaustive et excluant la lecture coréférentielle.

Les limites de cette étude sont soulignées par Jasinskaja elle-même :

D'abord, seules les relations de coréférence événementielle sont modélisées dans cette étude. Il ne s'agit que d'un sous-type particulier de l'*Élaboration*. Jasinskaja (2010) souligne de plus le manque de consensus qui existe concernant l'identité événementielle en général.

La deuxième critique avancée concerne un éventail d'autres relations de discours qui peuvent être inférées par défaut, i.e. sans la présence d'un marqueur du discours ou d'une marque intonative spécifique, mais qui n'impliquent pas la coréférence événementielle : *Explication*, *Évidence* et *Arrière-plan*.

4.4.2 Des études sur la signalisation de la relation d'*Élaboration*

4.4.2.1 SPRIK : le marqueur *indem*

Fabricius-Hansen et Behrens (2001) et Behrens et Fabricius-Hansen (2002) proposent une étude des marqueurs prototypiques de la relation d'*Élaboration* en allemand et en norvégien et leur traduction anglaise (*indem* (allemand), *ved å* (norvégien) et *by + ing* (anglais)) dans une approche de linguistique contrastive (Projet Språk i kontrast : SPRIK). Ce type d'étude est rendu possible grâce au corpus OMC (*Oslo Multilingual Corpus*) contenant une collection de textes originaux et traduits en anglais, français, norvégien et allemand. Selon elles, il existe un marqueur prototypique de la relation d'*Élaboration* en allemand, *indem*. Le marqueur est de plus associé à une liste de contraintes : les deux propositions reliées par le marqueur *indem* doivent avoir le même aspect, le même agent, la même extension spatio-temporelle. La présence de *indem*, si toutes les contraintes sont vérifiées, permet de signaler la relation d'*Élaboration* entre deux éventualités :

it (INDEM) demands that we understand S2 as describing the same eventuality as S1, or part of that event, in a more specific or less abstract manner.
(Fabricius-Hansen et Behrens, 2001, p. 10)

Le marqueur du discours *indem* a des équivalents de traduction en anglais avec la construction syntaxique *by + Ving* et en norvégien avec le marqueur du discours *ved å* :

- (20) (i) DaSS Leni selbst eine «Partikularistin» ist, beweist sie täglich, indem sie sämtliche Brötchenkrümel vom Teller aufliest und in den Mund steckt.
(ii) That leni herself is a «Particlist» is something she proves daily by gathering up all the bread crumbs from her plate and putting them into her mouth.
(iii) At Leni selv «Partikularist» beviser hun daglig ved å samle opp alle brødsmlene fra tallerkenene og putte dem i munnen.
(Fabricius-Hansen et Behrens, 2001, p. 6)

En français, le marqueur *indem* sera généralement traduit au moyen d'un gérondif (34 «indem» sur 43 sont traduits en français par le gérondif dans le corpus OMC :

- (21) (iv) Le fait qu'elle se considère comme une «Particulariste», Leni le prouve chaque jour en ramassant toutes les miettes de pain de son assiette pour les porter à sa bouche. (Notre traduction)

Mais nous verrons dans le chapitre 7 que la proposition gérondive (tout comme la proposition introduite par *indem*) n'est pas toujours une élaboration de la proposition principale. Nous nous appuierons pour notre étude sur les données du corpus OMC. Ce corpus fournit un matériel de recherche unique pour des études de linguistique contrastive. Bien que notre travail ne se situe pas dans une telle perspective, observer les traductions en français de *indem* (allemand) et *ved å* (norvégien) est une ouverture pour éclairer la relation d'Élaboration et son marquage en français. Nous souhaitons nous inspirer de la démarche de Fabricius-Hansen et Behrens (2001) et Behrens et Fabricius-Hansen (2002) pour nos études des marqueurs du discours afin de dégager leur sémantisme et de voir dans quelle mesure ceux-ci aident le lecteur à inférer une relation de discours particulière.

4.4.2.2 Les *Second-level Discourse Markers* de Siepmann (2005)

Dans un autre type d'analyse contrastive, Siepmann (2005) propose une étude basée sur corpus d'unités complexes appelées *multi-words units*. Son objectif est d'identifier et de décrire des unités lexicales qui assument des fonctions identiques ou similaires en anglais, français et allemand. Il procède en deux étapes : d'abord en faisant un inventaire des classes lexicales en considération, puis en catégorisant les membres de ces classes. L'auteur a pour hypothèse que les correspondances de ces items lexicaux, qu'il appelle *Second-Level Discourse Marker* (SLDM), peuvent se faire sur la base de la fonction qu'ils occupent dans les textes.

Bien que la RST ne tienne pas compte des marqueurs du discours dans les définitions des relations de discours, Siepmann (2005) travaille dans le cadre de cette théorie pour distinguer des catégories fonctionnelles pour le classement des SLDMs. Il distingue entre autres les *summarizers* et *concluders* qui constituent des indices la relation de *Résumé* de la RST, les *reformulators* et *resumers* qui constituent des indices de la relation de *Reformulation* et les *exemplifiers* pour la relation d'Élaboration. Cette liste de SLDMs constitue bien évidemment un point de départ à notre recherche de marqueurs de la relation d'Élaboration.

Nous remarquons toutefois l'hétérogénéité des listes fournies par Siepmann (2005). Concernant les exemplifieurs, il liste des marqueurs lexicaux (*par exemple*), mais également des impératifs (*prenez* (*,* *par exemple*)), des expressions nominales (*premier exemple*, *deux exemples*), des propositions infinitives (*pour prendre un exemple*), des constructions avec les verbes *illustrer*, *montrer*, *fournir*, *offrir*... (*Un exemple illustrera l'importance...*), et encore d'autres constructions non classées (*les exemples abondent...*, *au travers d'exemple comme celui-ci*).

Concernant notre objectif, cette liste pose les mêmes problèmes que les marqueurs du discours :

- Certains *Second-Level Discourse Markers* sont ambigus, on ne peut donc pas les considérer comme des marqueurs d'exemplification fiables. La liste contient la forme impérative du verbe *prendre*, *prenez*. Sans compter le fait que les verbes sont généralement exclus de la catégorie des marqueurs du discours, *Prenez* n'a pas toujours pour fonction d'introduire un exemple.

- (22) Pendant un temps, nous nous sommes bercés d'illusions sur l'économie japonaise, dont finalement des pans entiers sont encore au stade médiéval. **Prenez** l'agriculture : le bol de riz coûte dix fois plus cher qu'en Thaïlande. (Le Monde)
- (23) Le geste est impératif, le photographe exigeant. **Prenez** la pose, souriez, ordonne-t-il de la main.

Dans certains contextes, *Prenez* introduit un exemple comme en (22) mais ce n'est pas toujours le cas (23).

- Certaines expressions construites avec le mot «exemple» données comme marqueurs d'exemplification, comme ci-dessous «Tous ces exemples montrent...» n'introduisent pas une exemplification :

- (24) A Singapour, par exemple, l'augmentation de la prévalence de la myopie suit exactement l'augmentation de la durée de la scolarité. De 15,4 % de myopes chez les garçons n'ayant reçu aucune éducation scolaire, en 1992, la proportion passe à 65 % chez les diplômés de l'Université. Au Japon, on a noté une augmentation de 30 % des élèves myopes à l'âge de 17 ans entre 1984 et 1996, augmentation survenue après l'âge de 12 ans. A Taïwan enfin, 100 % des enfants des classes aisées - et donc éduquées - sont myopes. **Tous ces exemples** pris en Asie **montrent** une très forte prévalence de la myopie, vraisemblablement en partie d'origine génétique, mais qui augmente fortement du fait des études. (Wikipedia)

Dans ce cas, «Tous ces exemples montrent» permet d'introduire une conclusion obtenue à partir des faits exposés plus haut. L'expression «ces exemples» ne participe pas à la cohérence du texte mais plutôt à la cohésion.

- Enfin, il faudrait déterminer pour un certain nombre d'expressions de cette liste *considérez par exemple, l'autre exemple, le premier exemple, deux exemples* la contribution au niveau discursif des différents éléments qui les constituent. *Considérez par exemple* est formé à partir du marqueur du discours «par exemple» qui a par ailleurs été analysé comme un marqueur de la relation d'exemplification par Rossari (2000). *Le premier exemple* est formé à partir de l'adjectif numéral étudié par Schnedecker (2007) dans la mise en œuvre des relations de *Parallèle*, *Contraste* et *Élaboration*.

En plus des exemplifieurs, Siepmann (2005) fournit une liste de reformulateurs (*pour le dire en d'autres termes, j'ai cité précédemment, autrement dit, pour le dire autrement...*), d'énumérateurs (*pour commencer, enfin et surtout, inutile d'ajouter que...*), des résumeurs ou conclueurs (*une dernière remarque sur, nous terminerons en indiquant..., il nous reste à...*). Tous ces marqueurs sont considérés par Siepmann (2005) comme pouvant apparaître dans des segments reliés par la relation d'*Élaboration*.

Nous notons dans cette liste la présence d'opérateurs aspectuo-temporels (*nous terminerons*), qui ont par ailleurs été étudiés par Danlos (2005).

Concernant les accentueurs (*précisons que..., il faut bien être conscient que...*), les avertisseurs (*pour information, retenons que...*) et les marqueurs de clarification (*soyons bien clair, entendons*), nous ne voyons pas clairement la façon dont ceux-ci peuvent jouer un rôle dans la mise en œuvre de la relation d'*Élaboration*.

Enfin les introducteurs de topique (*abordons maintenant, passons à*) semblent effecti-

vement être de bons candidats pour introduire un topique qui sera ensuite élaboré dans le discours.

4.4.2.3 LEXCONN : liste de marqueurs lexicaux

Chercher à établir une corrélation entre marqueurs lexicaux et relations de discours est une tâche périlleuse. L'idée que cette corrélation existe est d'ailleurs une hypothèse très forte qui n'a jamais été démontrée. Bien évidemment on peut comprendre une telle démarche. Pour Roze *et al.* (2010), l'intérêt est essentiellement porté sur les enjeux d'une telle ressource pour le TAL. La ressource LEXCONN contient plus de 300 marqueurs lexicaux mais nous allons uniquement illustrer notre point de vue sur la ressource en nous appuyant sur la dizaine de marqueurs corrélés avec la relation d'Élaboration.

Parmi ceux-ci, *d'abord, premièrement, pour commencer, d'une part... d'autre part, d'un côté... d'un autre côté, déjà* sont considérés comme ambigus, marquant soit la relation d'Élaboration, soit la relation d'Explication. Cela signifie-t-il que les auteurs considèrent deux homonymes pour chacun de ces marqueurs : l'un spécifiant la relation d'Élaboration et l'autre spécifiant la relation d'Explication ? Nous verrons dans le chapitre 6 qu'une telle analyse pour *d'abord* et *dans un premier temps* ne convient pas et qu'il reste encore à démontrer si ce sont des marqueurs et si oui, de quelles relations.

Etablir des listes de marqueurs d'une relation de discours nécessite que les contextes d'apparition de ces marqueurs soient finement analysés.

4.4.2.4 Le projet ILF : Relations de cohérence et fonctionnement des anaphores

Pour terminer, nous allons présenter le projet ILF qui a donné lieu à la publication du numéro 19 de la revue *Journal of French Language Studies* consacré à la relation d'Élaboration. Ce numéro est le fruit du projet ILF : «Relations de cohérence et fonctionnement des anaphores» dans lequel ce projet de thèse est né. Plusieurs contributions de ce numéro thématique ont eu pour objectif de mettre au jour des indices qui permettraient de reconnaître la relation d'Élaboration. Deux articles étudient le rôle des adverbes *d'abord* (Bras et Le Draoulec, 2009) et *notamment* (Vergez-Couret, 2009) dans la mise en œuvre des relations de discours.

Dans le chapitre 5, nous présenterons nos analyses sur les adverbes paradigmatiques *notamment, particulièrement, en particulier, précisément*.

Les travaux sur les marqueurs de structuration du discours nous ont amenée à considérer les structures textuelles appelées structure énumérative. Ces structures permettent d'organiser le texte tant à un niveau global (organisation du texte dans sa globalité) qu'à un niveau plus local. Elles constituent un point de rencontre intéressant entre les approches descendantes (qui analysent le texte de façon sporadique en tentant de dégager les structures organisatrices du texte) et les approches ascendantes (qui analysent le texte de façon exhaustive en tentant de lier tous les segments du texte par des relations de discours). Notre hypothèse se base sur l'interdépendance des deux niveaux d'analyse : des structures

d'ordre textuel apportent des informations sur la façon dont les segments sont reliés au niveau sémantique (chapitre 6).

Les structures syntaxiques, comme le participe présent et le gérondif, ont été relevées comme jouant un rôle dans la mise en œuvre de la relation d'*Élaboration*. Nous analyserons la construction gérondive dans le chapitre 7.

Enfin, notre dernière piste de recherche concernant les indices de la signalisation de la relation d'*Élaboration* est de prendre en considération les relations lexicales qui s'établissent entre les prédicats et arguments en jeu lorsque la relation est vérifiée entre deux propositions ou deux phrases. Il s'agit d'un enjeu considérable dans ce travail de thèse. Il faut d'abord établir les types de relation lexicale qui jouent effectivement un rôle dans la mise en place et l'identification des relations de discours. Nous avons vu que la notion de sous-type lexical, basée sur la hiérarchie de type, nécessite que soient prises en compte d'autres informations d'ordre lexical ou des connaissances du monde. Et il s'agit également de trouver une ressource lexicale dans laquelle ces relations apparaissent afin d'envisager un repérage automatique de la relation (chapitre 10).

4.5 Bilan et positionnement

Identifier des indices et marqueurs du discours pour repérer les relations de discours permet d'une part de valider les structures de discours décrites dans les théories en mettant au jour des formes de surface leur correspondant et d'autre part de découvrir de nouvelles structures et d'affiner les connaissances sur des structures existantes.

Poursuivre cet objectif dans le cadre de la SDRT nous permet de disposer des outils fournis par la théorie pour accomplir cette étude systématique des indices et marqueurs de la signalisation du discours.

Enfin, identifier des indices de surface, repérables automatiquement, est un préalable pour toute application de TAL qui nécessite des informations sur la structure discursive des textes.

Suivant (Rossari, 2000), pour analyser un éventuel marqueur d'une relation, nous comparons d'une part la relation marquée et la relation non marquée (opération de suppression du marqueur) et d'autre part nous comparons le marqueur étudié avec d'autres marqueurs de la même famille (signalant la même relation de discours). L'objectif est de déterminer les fonctions discursives des marqueurs étudiés : sont-ils des producteurs de la relation ? des spécificateurs ? permettent-ils de bloquer l'inférence d'autres relations de discours ?

L'objectif est de déterminer pour chaque marqueur étudié :

- le rôle que celui-ci joue dans la mise en œuvre d'une relation de discours ;
- l'influence du contexte sur son interprétation ;
- et si son rôle est déterminé au niveau du contenu du constituant ou au niveau discursif.

Nos futures analyses se situent dans le cadre de la SDRT qui accorde une place primordiale aux facteurs linguistiques et contextuels contribuant à la représentation qu'un locuteur se fait de la structure du discours. Toutefois, ce modèle théorique s'est bâti autour de quelques exemples construits dont les analyses ont souvent été discutées. L'originalité

de notre travail est de proposer la confrontation de la théorie à des données attestées en s'appuyant sur les descriptions linguistiques les plus fines possibles. Pour accomplir cet objectif, nous allons donc nous appuyer sur les définitions d'*Élaboration* fournies dans (Asher et Lascarides, 2003) qui concernent uniquement les *Élaborations d'éventualité* (nous pouvons expliquer ce constat par le fait que la SDRT, à ses débuts, a puisé son inspiration dans les textes de type narratif. L'observation de textes de nature différente (tels que les textes descriptifs) a entraîné l'introduction d'une nouvelle relation de discours, *Élaboration d'entité* (Prévot *et al.*, 2009)). Pour notre analyse des élaborations, nous proposons trois types d'approches :

- Une approche de type sémasiologique, consistant à partir de marqueurs potentiels de la relation pour déterminer leur rôle dans la mise en oeuvre de la relation d'*Élaboration*. Parmi ces derniers, nous avons distingué des items lexicaux, des structures syntaxiques et des structures textuelles. Nous focaliserons notre attention sur un type de chaque : les adverbies paradigmatiques pour les items lexicaux (chapitre 5) ; les structures énumératives pour les structures textuelles dans lesquelles plusieurs types de marqueurs jouent un rôle dans la mise en oeuvre des relations de discours (chapitre 6) et le gérondif pour les structures syntaxiques (chapitre 7). Notre objectif est de fournir de nouvelles règles d'inférence qui viennent compléter celles déjà existantes sur la relation d'*Élaboration*.
- Une approche de type onomasiologique, consistant à partir de la relation pour en appréhender ses réalisations linguistiques et à valider les marqueurs potentiels mis au jour dans les chapitres précédents. Pour ce travail nous avons disposé à la fin de la thèse des données issues du corpus ANNODIS annoté en relations de discours (chapitre 9).
- Une approche de type TAL visant le repérage automatique de la relation d'*Élaboration* en combinant des indices ambigus de la relation mis au jour par l'approche sémasiologique et des indices de cohésion lexicale permettant de capturer les relations lexicales en jeu dans la mise en oeuvre de la relation (chapitre 10).

Deuxième partie

Analyse de marqueurs potentiels de l'*Élaboration* en SDRT

Dans l'état de l'art, nous avons présenté différentes approches de l'analyse du discours et du rôle des indices et marqueurs du discours dans ces approches. Dans cette seconde partie de la thèse, nous nous concentrons à établir les fonctions discursives de marqueurs potentiels dans la mise en œuvre de la relation d'*Élaboration*. Les marqueurs du discours sont centraux pour la reconnaissance de la structure discursive et des relations de discours. Ils facilitent et orientent l'interprétation et participent ainsi à la construction du sens du discours.

Dès lors que les études cherchent à repérer automatiquement la structure du discours, l'utilisation d'une ressource mettant en corrélation marqueurs du discours et relations de discours semble inévitable. Mais la construction d'une telle ressource se heurte à la complexité du rôle et de la place des marqueurs dans l'interprétation. Nous le savons bien, les marqueurs sont ambigus et pourtant de nombreux travaux fournissent des listes de relations de discours assorties de listes de marqueurs.

Dans cette deuxième partie, plutôt que de fournir la liste la plus exhaustive possible des marqueurs potentiels de la relation d'*Élaboration*, nous souhaitons nous concentrer sur une liste fermée de ces marqueurs et fournir les analyses les plus approfondies possibles de leur rôle dans la mise en place de la relation. Ces études ont trouvé leur source dans le cadre du projet ILF *Relations de cohérence et fonctionnement des anaphores*, consacré, comme nous l'avons déjà mentionné, majoritairement à une exploration approfondie de la relation de discours d'*Élaboration*. Un des objectifs fixés par ce projet est de déterminer s'il existe des marqueurs ou plus largement des indices de la relation.

Cette deuxième partie comporte trois chapitres. Dans le chapitre 5, nous analyserons les adverbes paradigmatiques *notamment*, *en particulier*, *particulièrement* et *précisément*. Leur rôle est, à première vue, d'apporter des éléments de détails. Nous verrons leur fonction dans la mise en œuvre de la relation d'*Élaboration*, que nous avons présentée comme s'établissant entre une information générale et une information plus spécifique portant sur une éventualité ou une entité.

Dans le chapitre 6, nous analyserons les structures énumératives. Celles-ci sont repérables via une configuration d'indices (amorces, marqueurs d'item...) et forment des «blocs» assumant une fonction rhétorique particulière. Nous verrons si les définitions de l'*Élaboration* dans la SDRT permettent de rendre compte des processus mis à l'œuvre dans l'interprétation de ces structures.

Enfin dans le chapitre 7, nous étudierons une structure syntaxique, le gérondif. Comme nous l'avons vu, les structures syntaxiques ont été moins étudiées du point de vue de la mise en œuvre des relations de discours. Nous verrons qu'il est nécessaire de passer par une analyse sémantique pour rendre compte des diverses interprétations possibles avec cette structure syntaxique.

Nous montrerons ainsi la complexité du rôle des marqueurs dans l'interprétation des discours et la mise en œuvre des relations de discours. Une analyse détaillée (sur de nombreux exemples attestés) pour chaque marqueur est requise avant de construire une ressource permettant l'automatisation du repérage de la structure du discours (*Cf.* Partie 3). Nos analyses porteront sur des données attestées issues d'une compilation d'articles de Wikipedia, Le Monde, l'Est Républicain et du corpus Géopo. Le modèle de la SDRT est parfaitement approprié pour accompagner cette attention très fine portée au détail des réalisations linguistiques aux niveaux syntaxique, lexical, sémantique et pragmatique

et à leurs interactions pour nous permettre de déterminer les contextes dans lesquels nos marqueurs corrèlent avec la mise en œuvre de la relation d'*Élaboration*.

Chapitre 5

Les adverbess paradigmatissants

*Cet adverbe notamment n'est pas du bel usage, il faut plutost dire nommément,
les meilleurs sont particulièrement, principalement, surtout...
(Mystères du Viel Testament, Vaugelas)*

Sommaire

5.1	Les adverbess paradigmatissants	130
5.1.1	Définitions	130
5.1.2	Position de l'adverbe et nature du paradigme	130
5.1.3	Une classe d'adverbess à part	132
5.1.4	Rôle de focalisation <i>vs.</i> rôle paradigmatissant	134
5.2	Rôle discursif des adverbess en position intraphrastique	138
5.2.1	<i>Notamment</i>	138
5.2.1.1	Paradigme préssupposé explicite	138
5.2.1.2	Paradigme préssupposé implicite	142
5.2.2	<i>Particulièrement</i>	146
5.2.3	<i>En particulier</i>	147
5.2.4	<i>Précisément</i>	148
5.2.5	Conclusion sur les emplois intraphrastiques	149
5.3	Rôle discursif des adverbess en position initiale détachée	150
5.3.1	Spécificité de la position initiale détachée	150
5.3.1.1	Emplois intraphrastiques <i>vs.</i> emplois à l'initiale détachée	150
5.3.1.2	Forme simple <i>vs.</i> forme comparative	151
5.3.2	<i>Notamment, en particulier</i>	152
5.3.3	<i>Plus particulièrement</i>	154
5.3.3.1	<i>plus particulièrement</i> paradigmatissant et conjonctif	155
5.3.3.2	Autres emplois de <i>plus particulièrement</i>	159
5.3.4	<i>Plus précisément</i>	162
5.3.5	Conclusion sur les emplois à l'initiale	165
5.4	Conclusion	166

Le point de départ de ce chapitre est une étude de l'adverbe *notamment* dans ses emplois intraphrastiques (Vergez-Couret, 2009), à partir des analyses de Nolke (1983) qui le classe

dans les adverbes paradigmatisants et de celles de Molinier et Lévrier (2000) qui le classent dans les adverbes focalisateurs. Nous proposons de compléter cette étude avec d'une part l'analyse de trois autres adverbes des mêmes classes - *particulièrement*, *précisément* et *en particulier* - et d'autre part celles des emplois en position initiale (dans leur forme simple et leur forme comparative le cas échéant) de ces quatre adverbes. Leur rôle est, à première vue, d'apporter des éléments de détails et/ou de précision. C'est en ce sens que nous nous intéressons à ces adverbes dans la mise en œuvre de la relation d'*Élaboration*.

Nous commençons par présenter les principales propriétés et caractéristiques des adverbes paradigmatisants (section 5.1). Puis nous présentons une analyse discursive des emplois intraphrastiques des quatre adverbes retenus (section 5.2). Nous terminons avec une analyse discursive de leurs emplois initiaux (section 5.3). Nous nous intéressons tout particulièrement à ces emplois à l'initial du fait que cette position est propice à conférer à l'adverbe un rôle de connexion.

5.1 Les adverbes paradigmatisants

5.1.1 Définitions

Les adverbes paradigmatisants - *notamment*, *particulièrement*, *précisément* et *en particulier* - se définissent syntaxiquement par leur grande mobilité au sein de la phrase. Ils sont très mobiles comme les adverbes de phrase. Mais, contrairement à ces derniers, un changement de position dans la phrase entraîne généralement deux interprétations différentes de l'énoncé. Sémantiquement, les adverbes paradigmatisants introduisent une présupposition entraînant que la phrase contenant l'adverbe soit mise en relation avec un paradigme d'autres phrases : «Il existe un paradigme de phrases qui intervient nécessairement dans l'interprétation de la phrase actuelle. À partir de la phrase actuelle dépourvue de l'adverbial, on obtiendra les phrases du dit paradigme en remplaçant le noyau par un autre membre de la même catégorie conceptuelle» (Nolke, 1983, p. 22). Les adverbes *notamment*, *en particulier*, *particulièrement* et *précisément* permettent de distinguer, à l'intérieur d'un paradigme de référence que la phrase dans son ensemble permet de définir, un sous-ensemble désigné par le constituant auprès duquel il figure.

5.1.2 Position de l'adverbe et nature du paradigme

Les adverbes paradigmatisants se caractérisent par leur grande mobilité au sein de la phrase. L'adverbe peut effectivement occuper les cinq positions décrites ci-dessous (Nolke, 1983, p. 17) :

- (1) 1. Paul 2. a 3. mangé 4. des gâteaux 5.

Les changements de position de l'adverbe paradigmatisant entraînent des changements de sens étant donné que ce dernier est toujours lié à un membre de la phrase avec lequel il se déplace dans la construction clivée.

- (2) C'est *notamment Paul* qui a mangé des gâteaux. (Position 1.)
C'est *Paul notamment* qui a mangé des gâteaux. (Position 2.)

C'est *notamment des gâteaux* que Paul a mangé. (Position 4.)

C'est *des gâteaux notamment* que Paul a mangé. (Position 5.)

Le membre de la phrase avec lequel l'adverbe se déplace constitue le pivot du paradigme.

Dans la position 1. (à l'initiale mais détaché) et 3., *notamment* ne focalise plus sur un élément de la phrase mais sur la phrase entière.

Concernant *notamment*, Molinier et Lévrier (2000, p. 281) signalent que celui-ci peut apparaître en position initiale détachée, position dans laquelle il occupe une fonction d'adverbe conjonctif :

- (3) Max a beaucoup d'activités. **Notamment**, il écrit un livre.

En tant que déclencheur de présupposition, nos adverbes, occupant les positions 1. et 3., ne peuvent pas introduire un discours et ils nécessitent un rattachement au contexte gauche immédiat.

Dans la position 1., *notamment* a le rôle d'un adverbe conjonctif et la phrase le précédent fait mention de l'ensemble des éventualités du paradigme comme illustré en (3) et (4).

- (4) Les enfants ont pris leur goûter. **Notamment**, Paul a mangé des gâteaux.

Dans ce cas, c'est l'éventualité «Paul a mangé des gâteaux» qui constitue le pivot du paradigme. Ce dernier est, de ce fait, constitué d'un ensemble d'éventualités, par exemple «Pierre a bu du jus d'orange», «Marie a dégusté des cerises du jardin»...

- (5) Les enfants ont pris leur goûter. **Notamment**, Paul a mangé des gâteaux et Pierre a bu un verre de jus d'orange.

Nous notons que *particulièrement* occupe difficilement cette position, contrairement à la forme comparative de celui-ci *plus particulièrement*.

Quand il apparaît en position 3., le paradigme est également constitué d'éventualités, comme vu précédemment, mais dans cette position, *notamment* ne semble pas jouer le même rôle de connection et ne nécessite pas que le paradigme d'éventualités soit mentionné dans le contexte gauche.

- (6) Ils <les délégués départementaux> sont désignés pour quatre ans par l'inspecteur d'académie. Ils peuvent **notamment** être consultés sur la convenance des projets de construction, d'aménagement et d'équipement des locaux (...) (L'Est républicain)

Le paradigme présupposé par *notamment* est constitué, en quelque sorte, de toutes les fonctions pouvant être attribuées aux délégués départementaux. Dans ce cas, *notamment* peut difficilement occuper la position initiale :

- (7) ?? Ils <les délégués départementaux> sont désignés pour quatre ans par l'inspecteur d'académie. **Notamment** ils peuvent être consultés sur la convenance des projets de construction, d'aménagement et d'équipement des locaux (...)

Au contraire, lorsque le paradigme est explicite, l'adverbe peut occuper la position 1. et la position 3. sans modification de sens.

- (8) En marge du festival, de multiples animations de plein air offrent un air de gaieté à la ville. Les «arts Bladabloum clowns» ont **notamment** présenté un drôle de spectacle, alliant le jonglage, le feu et les parades. (L'Est Républicain)
- (9) En marge du festival, de multiples animations de plein air offrent un air de gaieté à la ville. **Notamment** les «arts Bladabloum clowns» ont présenté un drôle de spectacle, alliant le jonglage, le feu et les parades.

Bilan

De ces manipulations, nous retenons deux faits importants :

- 1. les adverbes paradigmatiques ont la capacité d'isoler un élément ou un sous-ensemble d'un paradigme, et
- 2. en position 1. (à l'initiale détachés) et 3., ce paradigme est constitué d'éventualités.

5.1.3 Une classe d'adverbes à part

Les caractéristiques que nous venons de présenter sont une explication possible aux difficultés rencontrées lorsque l'on cherche à classer ces adverbes dans un système binaire d'adverbes intégrés à la proposition et d'adverbes de phrase.

Selon Molinier et Lévrier (2000, p. 44-46), les adverbes de phrase sont définis par la conjonction des deux propriétés suivantes :

- 1. Possibilité de figurer en position détachée en tête de phrase négative.
- 2. Impossibilité d'extraction dans *c'est... que*

tandis que les adverbes intégrés à la proposition sont définis par la vérification de l'une et/ou l'autre des deux propriétés suivantes :

- 1. Impossibilité de figurer en position détachée en tête de phrase négative.
- 2. Possibilité d'extraction dans *c'est... que*

Pour définir la classe des adverbes focalisateurs, Molinier et Lévrier (2000, p. 273) proposent trois propriétés adaptées des tests permettant de classer adverbes de phrase et adverbes intégrés à la proposition.

- 1. Impossibilité de figurer en position détachée en tête de phrase négative.
- 2. Impossibilité d'extraction dans *c'est... que*
- 3. Possibilité d'extraction dans *c'est... que* en compagnie d'un constituant majeur de la phrase (SN, SP)

Les emplois intraphrastiques de nos adverbes sont ainsi bien décrits mais cette définition ne permet pas de décrire les emplois initiaux de ceux-ci. Nous allons illustrer ces deux points en appliquant les propriétés des adverbes focalisateurs avec nos adverbes *notamment*, *en particulier*, *particulièrement* et *précisément* dans la phrase test de Molinier et Lévrier (2000, p. 273).

(10)

$$\text{Max boit} \left\{ \begin{array}{l} \text{notamment} \\ \text{en particulier} \\ \text{particulièrement} \\ \text{précisément} \end{array} \right. \text{ du vin.}$$

Propriété 1. Impossibilité de figurer en position détachée en tête de phrase négative

Cette transformation est possible mais elle implique un changement dans la portée de la focalisation.

(11)

$$\left\{ \begin{array}{l} \text{Notamment,} \\ \text{En particulier,} \end{array} \right. \text{ Max ne boit pas de vin.}$$

Dans une telle position, *notamment* et *en particulier* focalisent sur la phrase entière et fonctionnent ainsi comme des adverbes de phrase conjonctifs focalisateurs. Ils ne peuvent pas focaliser sur le SN «vin», comme ils le font en (10).

Concernant *particulièrement*, l'emploi initial d'adverbe de phrase conjonctif focalisateur est mieux supporté par la forme comparative que par la forme simple.

(12)

$$\left\{ \begin{array}{l} \text{? Particulièrement,} \\ \text{Plus particulièrement} \end{array} \right. \text{ Max ne boit pas de vin.}$$

Enfin, *précisément* n'a plus dans ce cas l'emploi de focalisateur mais celui d'un adverbe de phrase conjonctif transitionnel au sens de Molinier et Lévrier (2000, p. 57) qui a ici pour fonction de marquer la coïncidence entre deux situations ou deux faits. Il pourrait être remplacé par *justement*. En revanche, à la forme comparative, nous notons que *plus précisément* a un emploi d'adverbe de phrase conjonctif focalisateur de l'ensemble de la phrase.

(13)

$$\left\{ \begin{array}{l} \text{Précisément,} \\ \text{Plus précisément,} \end{array} \right. \text{ Max ne boit pas de vin.}$$

Propriété 2. Impossibilité d'extraction dans *c'est... que*

Ils ne peuvent être extraits seuls dans *c'est... que* :

(14)

$$\text{C'est} \left\{ \begin{array}{l} \text{*notamment} \\ \text{*en particulier} \\ \text{*particulièrement} \\ \text{*précisément} \end{array} \right. \text{ que Max boit du vin.}$$

Propriété 3. Possibilité d'extraction en *c'est... que* en compagnie d'un constituant majeur de la phrase

Ils peuvent être extraits dans *C'est... que* en compagnie du constituant qu'ils accompagnent, ici le syntagme nominal «du vin» :

(15)

$$C'est \left\{ \begin{array}{l} \text{notamment} \\ \text{en particulier} \\ \text{particulièrement} \\ \text{précisément} \end{array} \right. \text{ du vin que Max boit.}$$

Ce test ne peut être appliqué aux emplois de *notamment* en position 1. (à l'initiale détaché) et en position 3. étant donné que la phrase entière ne peut être focalisée dans la clivée.

Bilan

Les définitions de Molinier et Lévrier (2000) s'adaptent bien aux emplois intraphrastiques de nos quatre adverbes. Ainsi les focalisateurs sont-ils considérés comme des adverbes intégrés à la proposition bien que ceux-ci nécessitent d'adopter une version adaptée de l'extraction dans *c'est... que*. Mais ces définitions ne s'appliquent pas aux emplois initiaux des adverbes.

Du point de vue sémantique que nous adoptons, nous retenons la capacité de ces adverbes à focaliser soit sur un élément de la phrase, soit sur la phrase toute entière.

C'est cette fonction focalisatrice et paradigmatitante des adverbes *notamment*, *en particulier*, *particulièrement* et *précisément* que nous souhaitons analyser en lien avec la mise en œuvre de la relation d'*Élaboration*.

5.1.4 Rôle de focalisation *vs.* rôle paradigmatissant

Molinier et Lévrier (2000) soulignent que les adverbes *notamment*, *en particulier*, *particulièrement* et *précisément* jouent un rôle de focalisation ou de contraste. Ces adverbes permettent de distinguer, à l'intérieur d'un paradigme de référence que la phrase dans son ensemble permet de définir, un sous-ensemble désigné par le constituant auprès duquel il figure.

Ce rôle d'extraction d'un ou de plusieurs éléments à l'intérieur d'un paradigme de référence a été considéré pour *notamment* dans plusieurs études. Ces études se distinguent toutefois par les types d'ensembles mis en relation par l'adverbe.

Notamment et la relation d'hypéronymie

Borillo (1996) décrit *notamment* comme un marqueur potentiel de la relation d'hypéronymie dans des énoncés génériques, comme dans l'exemple suivant :

(16) Chez les insectes, **notamment** chez les abeilles [...]

L'adverbe *notamment* met en relation dans le discours deux SN dont les noms sont les lexèmes considérés par la relation d'hypéronymie : l'hypéronyme «insecte» et l'hyponyme «abeille». Plus généralement, le patron <X, notamment Y> permet l'identification de deux lexèmes concernés par la relation d'hyperonymie, avec X comme hypéronyme et Y comme hyponyme. Borillo (1996) propose également les patrons <X et plus particulièrement Y>, <X, en particulier Y>, <X comme par exemple Y> ou <X, particulièrement Y> pour le repérage de l'hypéronymie en discours spécialisé.

- (17) Chez les insectes, **(et) (plus) particulièrement** chez les abeilles [...]
 (18) Chez les insectes, **en particulier** chez les abeilles [...]

Notamment et la relation de méronymie

Crévenat-Werner (2003) analyse *notamment* comme un marqueur de la relation de méronymie (ou partie à tout) entre deux référents extra-linguistiques, comme dans l'exemple suivant :

- (19) J'ai cueilli des fruits, **notamment** des poires que j'apporterai pour ton gâteau.
 (Crévenat-Werner, 2003, p. 171)

La relation de partie à tout (de type *sous-collection* : *collection* dans la terminologie de Aurnague et Vieu (1993)) relie, dans l'exemple ci-dessus, deux référents extra-linguistiques (dans l'ensemble des référents «fruits» sont inclus les référents «poires»). Cette relation de méronymie est construite sur une relation d'hypéronymie s'établissant entre les lexèmes mis en relation (avec l'hypéronyme «fruit» et l'hyponyme «poire»), comme dans l'exemple (16) de Borillo (1996). En (19), *particulièrement*, *en particulier* et *précisément* peuvent se substituer à *notamment* :

- (20)
- | | | |
|--------------------------|---|--|
| J'ai cueilli des fruits, | { | en particulier
particulièrement des poires...
précisément |
|--------------------------|---|--|

Ces substitutions entraînent des changements de sens majeurs au niveau de la relation établie entre les référents extra-linguistiques. Avec *particulièrement* et *en particulier*, la relation de méronymie entre les deux référents est conservée (les «poires» sont incluses dans l'ensemble des «fruits») mais cet ensemble de référents désigné par *particulièrement* est considéré comme le plus important en quantité ou en qualité dans l'ensemble des référents «fruits». Avec *précisément*, l'interprétation méronymique des deux référents est exclue. L'apposition apporte une redescription du référent «fruits» en employant un hyponyme de celui-ci, «poires». Du point de vue référentiel, les deux ensembles désignés par «fruits» et «poires» renvoient aux mêmes référents extra-linguistiques.

Examinons un autre exemple de marquage de relation de méronymie emprunté à Crévenat-Werner (2003).

- (21) J'ai mal dessiné la pomme, **notamment** les pépins. (Crévenat-Werner, 2003, p. 172)

La relation de meronymie de type *composant : assemblage* dans la terminologie de Aurnague et Vieu (1993) existe, ici, entre les lexèmes «pomme» et «pépin» mis en relation par *notamment* (la pomme est composée de pépins). Il s'ensuit que les référents dénotés par ces lexèmes sont dans une relation d'inclusion spatiale dans le monde (les pépins sont inclus dans la pomme).

Dans ce type d'énoncé, il est également possible d'insérer les adverbes *particulièrement*, *en particulier* et *précisément* :

- (22)
- J'ai mal dessiné la pomme, $\left\{ \begin{array}{l} \text{en particulier} \\ \text{particulièrement} \\ \text{précisément} \end{array} \right.$ les pépins.

Il s'ensuit également un changement de sens. Avec *particulièrement* et *en particulier*, dans l'ensemble du dessin de la pomme, les «pépins» vérifient à un plus haut degré la propriété d'avoir été «mal dessinés». Avec *précisément*, la pomme est mal dessinée car les «pépins», et seulement les pépins de celle-ci, ont été mal dessinés.

Nous pouvons résumer les analyses de Crévenat-Werner (2003) mises en perspective avec nos manipulations en (19) (20) (21), (22).

- L'élément attaché à *notamment* est choisi par le scripteur ou locuteur comme étant le plus important du paradigme, autrement dit le plus *notable*, comme son nom l'indique. Un élément est choisi dans un paradigme mais les éléments restants de ce paradigme ne sont pas hiérarchisés entre eux.

- Les éléments attachés aux adverbes *particulièrement* et *en particulier* vérifient à un degré plus élevé la propriété décrite par la phrase d'accueil. Le paradigme établi par *particulièrement*, contrairement à celui établi par *notamment*, est construit sur une propriété gradable, permettant ainsi une continuation avec *en second lieu*.

- (23) J'ai cueilli des fruits, **particulièrement** des poires, **en second lieu** des pommes.

- (24) J'ai cueilli des fruits, **en particulier** des poires, **en second lieu** des pommes.

La même continuation avec *notamment* semble plus difficile (Crévenat-Werner, 2003, p. 180) tandis qu'une continuation plus neutre avec un marqueur comme *également* est possible :

- (25) ? J'ai cueilli des fruits, **notamment** des poires, **en deuxième lieu** des pommes.

- (26) J'ai cueilli des fruits, **notamment** des poires, mais **également** des pommes.

- *Précisément* semble jouir d'un statut particulier. Il correspond plus à la définition du focalisateur qu'à celle du paradigmatissant. L'élément attaché à l'adverbe constitue une redescription en des termes plus précis (en utilisant un hyponyme ou une partie du tout). *Précisément* ne fonctionne pas toujours comme un adverbe paradigmatissant dans le sens où il n'y a pas nécessairement construction d'un paradigme de phrases intervenant

dans l'interprétation de la phrase globale. La continuation avec *en second lieu* n'est pas pertinente lorsque *précisément* introduit une redescription :

(27) *J'ai cueilli des fruits, **précisément** des poires, **en second lieu** des pommes.

Les adverbess *notamment*, *en particulier* et *particulièrement* sont des adverbess paradigmatissants au sens de Nolke (1983). L'emploi de ces adverbess présuppose l'existence d'un paradigme qui intervient nécessairement dans l'interprétation. Nous illustrons ce point avec l'exemple suivant :

(28) Parvati parle **notamment** l'hindi.

Imaginons un contexte dans lequel Parvati parle trois langues. Les phrases du paradigme seraient {Parvati parle l'hindi ; Parvati parle le Tamoul ; Parvati parle le Sanskrit}. Dans ce contexte, les exemples suivants sont irrecevables :

(29)

* Parvati parle trois langues, $\left\{ \begin{array}{l} \text{notamment} \\ \text{en particulier} \\ \text{particulièrement} \end{array} \right.$ l'hindi, le tamoul et le sanskrit.

(30) * Parvati parle **précisément** l'hindi.

L'exemple (29) montre que les emplois de *notamment*, *en particulier* et *particulièrement* sont incompatibles avec l'introduction de tous les éléments de l'ensemble, prouvant ainsi leur fonction paradigmatissante. L'exemple (30) montre l'incapacité de *précisément* à sélectionner un élément d'un ensemble dans ce contexte.

La fonction paradigmatissante des adverbess *notamment* et *particulièrement* et celle non paradigmatissante de *précisément* est également mise en évidence par les exemples suivants dans lesquels la phrase d'accueil décrit une éventualité indivisible (Crévenat-Werner, 2003, p. 182) :

(31)

* Je suis née $\left\{ \begin{array}{l} \text{notamment,} \\ \text{en particulier} \\ \text{particulièrement} \end{array} \right.$ le neuf août.

(32) Je suis née **précisément** le 9 août.

Dans (31), il n'existe pas d'éventualité que *notamment*, *en particulier* ou *particulièrement* puisse isoler d'un ensemble. Ces adverbess ne peuvent donc pas apparaître dans ce contexte tandis que *précisément* peut le faire.

Bilan

Les fonctions focalisatrices et paradigmatissantes sont des opérations qui sont le signe d'une «idée regardante portée par l'énonciateur sur tout ou partie de son énoncé» (Guimier, 1996, p. 103). Ils sont qualifiés par ce dernier d'*exophrastique*.

Les adverbes paradigmatissants sont considérés comme portant sur le dire :

- i. parce qu'ils opèrent une focalisation sur un élément qui est isolé afin d'être mis en évidence (aspect mis en évidence par Molinier et Lévrier (2000) avec les adverbes focalisateurs) ;
- ii. parce qu'ils signalent un non-dit, correspondant à l'ensemble des éléments du paradigme auquel l'adverbe renvoie (aspect mis en évidence par Nolke (1983) avec les adverbes paradigmatissants).

Les adverbes paradigmatissants sont des objets pertinents pour l'étude de la relation d'*Élaboration*. *Notamment*, *en particulier* et *particulièrement* permettent d'introduire des détails et/ou précisions en sélectionnant un élément *notable* ou *particulier* d'un paradigme présupposé par leur emploi. *Précisément* a un statut particulier. Il permet d'introduire une redescription, qui entre également dans le champ de l'*Élaboration*.

Nous allons commencer nos analyses sur les adverbes *notamment*, *particulièrement*, *en particulier* et *précisément* et leurs rôles discursifs en considérant leurs emplois intraphrastiques.

5.2 Rôle discursif des adverbes en position intraphrastique

Dans le cadre de cette étude sur la recherche de marqueurs et indices de la relation d'*Élaboration*, nous nous focalisons uniquement sur des cas où les adverbes paradigmatissants isolent des éventualités d'un paradigme d'éventualités, i.e. lorsque ceux-ci sont à l'intérieur de la phrase, occupant la position 3. Nous présentons, d'abord, une analyse détaillée de l'adverbe *notamment*. Puis nous analysons dans les mêmes contextes les trois autres adverbes *en particulier*, *particulièrement* et *précisément* pour dégager les similitudes et différences de fonctionnement.

5.2.1 *Notamment*

L'adverbe *notamment* a fait l'objet d'une première étude au cours de notre thèse (Vergez-Couret, 2009) dans laquelle nous avons montré que le rôle de l'adverbe dans la mise en oeuvre de la relation d'*Élaboration* diffère selon que le paradigme présupposé est explicite ou non.

5.2.1.1 Paradigme présupposé explicite

Nous avons vu précédemment que les adverbes paradigmatissants, dans la position 3., isolent une ou des éventualités d'un paradigme d'éventualités comme nous pouvons le voir dans l'exemple (33)

- (33) [Des peines de six mois à huit ans de prison ferme ont été requises, mardi 29 janvier, à l'encontre des seize militants présumés du mouvement nationaliste basque Iparretarrak, jugés devant la seizième chambre du tribunal correctionnel de Paris pour «association de malfaiteurs»]_a [Le substitut du procureur de la République, Mme Irène Stoller, a **notamment** réclamé une peine de huit ans d'emprisonnement

contre le chef présumé du groupe Philippe Bidart, ainsi que contre Jean-Gabriel Mouesca et Joseph Etchebeste.]]_b (*Le Monde*, 31 janvier 1991)

De la phrase (b) contenant l'adverbe paradigmatissant, trois éventualités sont isolées : {Le juge Irène Stoller a réclamé une peine de huit ans contre Philippe Bidart ; Le juge Irène Stoller a réclamé une peine de huit ans contre Jean-Gabriel Mouesca ; Le juge Irène Stoller a réclamé une peine de huit ans contre Joseph Etchebeste}. Le paradigme présupposé est constitué de l'ensemble des éventualités que l'on peut résumer ainsi : «**Des juges** ont réclamé **des peines** à l'encontre **de seize nationalistes basques**».

En (33), la phrase précédant celle contenant *notamment* explicite ce paradigme. Du point de vue des ensembles de référence établis par les deux phrases, l'ensemble d'éventualité isolé par *notamment* est plus précis que celui défini par la phrase précédente.

Dans le cadre d'une analyse SDRT, le prédicat *Subtype_D* est vérifié entre les types des éventualités décrites en (a) et (b). Le type de l'éventualité décrite en (b) «Mme Irène Stoller a réclamé une peine de huit ans d'emprisonnement» est une sous-partie du type de l'éventualité décrite en (a) «Des peines de six mois à huit ans de prison ferme ont été requises». Nous inférons donc *Élaboration* entre π_a et π_b . Dans cette analyse, la relation d'*Élaboration* est ainsi inférée, en l'absence de prise en compte de l'adverbe *notamment*.

Bien qu'il n'ait pas ici un rôle nécessaire, *notamment* semble fonctionner comme un appui facilitant l'inférence de la relation, par le découpage qu'il opère d'un sous-ensemble de peines à l'encontre de trois personnes.

Les exemples (34) et (35) illustrent d'autres cas dans lesquels le paradigme présupposé par *notamment* est précédemment donné dans le discours, mais nous allons voir que, dans ces deux exemples, la suppression de *notamment* n'est pas sans effet :

- (34) [En septembre 1985, un an et demi après le début de son mandat de maire, Bucaram commit une succession d'erreurs,]]_a [**notamment** en accusant divers hommes politiques dans l'implication de meurtres mal élucidés.]]_b (Wikipedia)

En (34), la construction gérondive met en relation les deux éventualités «commettre des erreurs» et «accuser divers hommes». Le prédicat *Subtype_D* peut être vérifié entre les deux types d'éventualité mais le lien est ici très distendu : beaucoup d'éventualités peuvent être qualifiées d'erreur. La construction gérondive joue un rôle important ici dans la mise en œuvre de la relation d'*Élaboration*, comme nous le verrons dans le chapitre 7.

De même dans l'exemple (35) :

- (35) [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]]_a [recevant **notamment** le représentant du gouvernement russe en Tchétchénie, Nikolaï Kochman.]]_b (Libération.fr, le 04 janvier 2000)

La relation d'*Élaboration* est inférée grâce au prédicat *Subevent_D* vérifié entre «consacrer sa journée à s'occuper du dossier» et «recevoir le représentant du gouvernement russe» sur la base du participe présent et de connaissances du monde. *Notamment* sélectionne une des activités que Vladimir Poutine a menée au cours de sa journée à s'occuper du dossier de la guerre en Tchétchénie.

Dans les deux cas présentés ci-dessus, la relation d'*Élaboration* est vérifiée en l'absence de prise en compte de l'adverbe *notamment*. Toutefois nous notons que sa suppression va entraîner un changement de sens :

- (36) En septembre 1985, un an et demi après le début de son mandat de maire, Bucaram commit une succession d'erreurs, en accusant divers hommes politiques dans l'implication de meurtres mal élucidés.
- (37) [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]_a [recevant le représentant du gouvernement russe en Tchétchénie, Nikolai Kochman.]_b

Le changement d'interprétation s'effectue sur le plan temporel. Dans les exemples (34) et (35), les éventualités décrites en (a) et (b) sont spatio-temporellement dans une relation d'inclusion stricte : «accuser divers hommes politiques» renvoie à un sous-ensemble des «erreurs commises par Bukara» et «rencontrer Nikolai Kochman» a été une des «activités de la journée de Vladimir Poutine». En revanche, dans les exemples (36) et (37), avec la suppression de *notamment*, il est alors nécessaire en (36) de considérer que la pluralité d'éventualités dénotée par «commettre des erreurs» est égale à celle dénotée par «accuser divers hommes politiques». Cela explique la bizarrerie en (38) où une éventualité unique est dénotée par «accuser le meneur de l'opposition».

- (38) ?? Bucaram commit une succession d'erreurs, en accusant le meneur de l'opposition dans l'implication de meurtres mal élucidés.

En (37), il n'est pas nécessaire mais seulement possible d'interpréter les deux événements comme étant spatio-temporellement en co-extension.

Nous retenons de ces manipulations que sur le plan spatio-temporel, la suppression de *notamment* entraîne le passage d'une inclusion stricte à une possible co-extension ou à une égalité spatio-temporelle des ensembles dénotés.

Dans le cadre de la SDRT, les contraintes spatio-temporelles d'*Élaboration* retenues sont l'inclusion ou la concomitance spatio-temporelle des segments reliés par la relation (pour traiter à la fois des cas d'*Élaboration* et de *Reformulation* Cf. chapitre 3). Toutefois, la présence de *notamment* est incompatible avec la co-extension spatio-temporelle. Nous avons dans les cas d'*Élaboration* avec *notamment* une inclusion spatio-temporelle stricte entre les éventualités qui pourra être inférée grâce à l'axiome 5.1.

Axiome 5.1 $\text{Élaboration}(\alpha, \beta) \wedge [\text{notamment}](\beta) \rightarrow \neg(e_\alpha \equiv_{st} e_\beta)$

Cet axiome permet de restreindre la relation d'inclusion inférée grâce à l'axiome des effets sémantiques d'*Élaboration* à une inclusion stricte en supprimant l'inférence de l'équivalence.

Si deux constituants π_α et π_β sont reliés par *Élaboration* et que le segment β contient *notamment*, alors l'éventualité décrite en π_β est spatio-temporellement strictement incluse dans l'éventualité décrite en π_α .

Pour poursuivre nos investigations sur le rôle discursif de l'adverbe *notamment*, nous allons considérer le discours suivant dans lequel est mis en oeuvre un cas typique de la relation d'*Élaboration* : une éventualité complexe décrite dans le segment élaboré est ensuite décrite en plusieurs phases dans les segments élaborants.

- (39) [Sans arme, sans argent, ils durent se débrouiller,]_a [volant la voiture d'un surveillant]_b,
[échouant au bord d'un marais,]_c, [passant leur première nuit de cavale dans un
gabion au bord de l'eau.]_d (Le Monde)

On peut d'abord inférer *Élaboration* entre le constituant π_a et le constituant π_b grâce au prédicat *Subtype_D* qui s'établit entre le type de l'éventualité «devoir se débrouiller» et le type de l'éventualité «voler une voiture» en s'appuyant sur des connaissances du monde et le participe présent. Puis nous inférons une relation de *Narration* entre π_b et π_c , puis entre π_c et π_d , également grâce à des connaissances du monde sur le fait que ces trois éventualités se suivent logiquement dans une séquence narrative. Les constituants π_b , π_c et π_d sont reliés au moyen de la relation coordonnante *Narration*, ils forment ainsi un constituant complexe, noté π_A . Le premier constituant π_b de ce constituant est, quant à lui, attaché au segment π_a par la relation d'*Élaboration*. Il s'ensuit, suivant le principe inverse de la distributivité (Cf. chapitre 3), que le segment complexe π_A élabore π_a . Les segments π_b , π_c et π_d expriment, avec plus de détails, la totalité des actions entreprises décrites par l'éventualité en π_a «sans arme, sans argent, ils durent se débrouiller». Nous pouvons transformer cet exemple en y insérant l'adverbe *notamment*, qui introduirait l'élément de la liste que nous souhaitons mettre en valeur :

- (40) [Sans arme, sans argent, ils durent se débrouiller,]_a [volant **notamment** la voiture
d'un surveillant.]_b

Dans les exemples (39) et (40), le constituant π_a joue le rôle de topique de discours dans la structure. Il est élaboré en (39) par la description des actions de débrouillardise entreprises par les protagonistes, et en (40) par la description d'une action illustrant la débrouillardise des protagonistes. Selon les termes de Nolke (1983), et tel que nous l'avons décrit dans la section 5.1.2, *notamment* présuppose, en (40), un paradigme d'actions de débrouillardise et isole une de ces actions. L'éventualité isolée par *notamment* est mise en valeur, considérée comme la plus notable par le scripteur.

L'emploi de *notamment*, isolant une éventualité, n'empêche pas que d'autres éventualités du paradigme soient citées dans le discours.

- (41) [Sans arme, sans argent, ils durent se débrouiller,]_a [volant **notamment** la voiture
d'un surveillant.]_b [Ils durent également passer leur première nuit de cavale dans
un gabion au bord de l'eau]_c.

Dans cet exemple, le constituant π_c introduit une autre action de débrouillardise. L'idée d'addition est véhiculée par l'adverbe *également* et le parallélisme syntaxique. Nous inférons la relation de *Narration* entre π_b et π_c , formant ainsi le constituant complexe π_A . Suivant le principe inverse de distributivité, il est alors possible d'inférer *Élaboration* (π_a , π_A).

Nous allons, à présent, analyser des discours dans lesquels le paradigme présupposé n'est pas explicite.

5.2.1.2 Paradigme présupposé implicite

Dans la construction avec *notamment* intraphrastique, le paradigme présupposé n'est pas nécessairement explicite.

Dans le premier exemple que nous allons considérer (42), la présence d'un autre marqueur du discours permet l'inférence d'une autre relation de discours qu'*Élaboration*.

- (42) [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [notamment à cause de scènes très crues entre Marlon Brando et Maria Schneider.]_b (Wikipedia)

Le marqueur du discours à *cause de* indique que le constituant π_b se rattache au segment π_a par la relation d'*Explication*.

Notamment se situe en position initiale du segment expliquant. Quelle est alors la nature du paradigme présupposé par l'emploi de l'adverbe ? L'ensemble présupposé est constitué, nous semble-t-il, des causes pour lesquelles le dernier Tango à Paris fait scandale. *Notamment* fait intervenir des éléments liés au contexte, c'est-à-dire des éventualités qui entretiennent, dans les connaissances du monde, un rapport de cause à effet avec l'éventualité «le film fait scandale». De ce fait, *notamment*, dans ce cas, semble davantage jouer un rôle sémantique au niveau du discours qu'au niveau de la phrase. Il porte sur «à cause de» : «la cause en est notamment...», «l'une des causes en est...». Il permet de spécifier qu'il y a d'autres causes, i.e. que l'éventualité «comprendre des scènes très crues» n'est pas la seule éventualité à entretenir un lien causal avec «faire scandale». Nous proposons de rendre compte de l'ensemble présupposé par *notamment* au moyen d'un constituant implicite, qui ne correspond donc pas à un segment dans le texte, mais qui intervient dans l'interprétation du discours. Ce constituant est noté π_1 dans la figure 5.1. Le constituant implicite joue le rôle de topique discursif et il est élaboré par le constituant π_b qui donne une des causes du scandale. Les tirets représentent les éléments implicites présupposés par *notamment* qui sont potentiellement coordonnés entre eux dans la structure du discours.

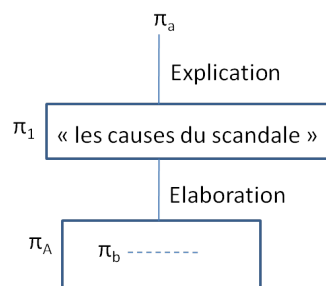


FIG. 5.1 – SDRS de (42) avec représentation du topique implicite noté π_1

Cette proposition entraîne la multiplication des topiques discursifs dans la structure, la rendant ainsi plus complexe. Il faut donc s'assurer que prendre en compte l'ensemble

présupposé est bien nécessaire au niveau de la représentation discursive. D'après Nolke (1983), il existe un paradigme d'éléments présupposés qui intervient nécessairement dans l'interprétation du discours. Suivant Guimier (1996), les adverbes paradigmatiques ont une portée sur le «dire». Ils sont le signe d'une série d'opérations dont on n'a pas à rendre compte au niveau du contenu propositionnel (le «dit») mais plutôt au niveau de la structure du discours. Plus précisément, les adverbes paradigmatiques signalent un «non-dit» dont on peut rendre compte au moyen d'un topique de discours implicite. Ces deux auteurs apportent, de notre point de vue, des arguments forts en faveur de la création du topique implicite composé du paradigme d'éventualités présupposés par *notamment*.

Cependant, si un constituant se trouve ainsi introduit dans la structure du discours sur la frontière droite, cela implique qu'il est disponible pour la poursuite du discours. Comme on le voit en (43), un attachement à ce topique implicite semble pour le moins difficile, voire impossible :

- (43) ?? [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [notamment à cause de scènes très crues entre Marlon Brando et Maria Schneider,]_b [Ces causes ont poussé les autorités à interdire le film aux moins de 16 ans.]_c

Concernant la poursuite du discours, nous voyons avec l'exemple suivant qu'il est possible de poursuivre le discours en faisant référence à une autre cause du scandale (tout comme l'on pouvait poursuivre les actions de débrouillardise de nos protagonistes en (41)) :

- (44) [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [notamment à cause de scènes très crues entre Marlon Brando et Maria Schneider,]_b [et également à cause des dialogues osés et des références machistes.]_c

Dans ce cas, le constituant π_c s'attache au constituant π_b par la relation de *Continuation*. Le constituant π_c apporte de nouvelles raisons du scandale, poursuivant ainsi la même fonction rhétorique que le constituant π_b . La contrainte de topique spécifie que deux constituants coordonnés forment un constituant complexe et doivent partager un topique commun, ici «les raisons du scandale».

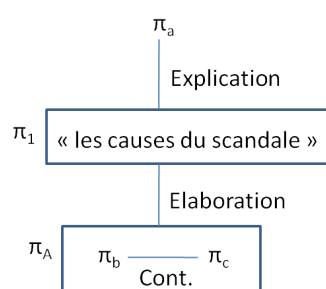


FIG. 5.2 – SDRS de (44) avec représentation du topique implicite noté π_1

La poursuite en «Ces raisons...» ne pose plus problème dans ce cas :

- (45) [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [notamment à

cause de scènes très crues entre Marlon Brando et Maria Schneider.]_b [Les dialogues osés et les références machistes sont d'autres causes d'indignation.]_c [Ces raisons ont poussé les autorités à interdire le film aux moins de 16 ans.]_d

Ces manipulations indiquent que *notamment* ne peut à lui seul justifier l'introduction d'un topique implicite dans la structure du discours, étant donné que ce topique ne paraît pas disponible pour la poursuite du discours. En revanche, dès qu'un autre élément du paradigme est introduit dans la structure, celui-ci est attaché au segment contenant *notamment* par une relation de coordination *Continuation* et entraîne l'introduction d'un topique implicite qui se trouve alors disponible pour la poursuite du discours. Ce topique implicite n'a toutefois pas le même statut que celui proposé précédemment. Dominant un constituant complexe, celui-ci a plus de contenu.

Le rôle de *notamment* serait donc exprimé au niveau du contenu du constituant, spécifiant que l'explication introduite «notamment à cause de» n'est qu'une explication partielle, Cf. Figure 5.3. Nous ne rentrerons pas dans le détail de la façon dont ce traitement est opéré, par exemple en s'appuyant sur les travaux de Rooth (1992), étant donné que cela nous éloigne de notre objectif d'analyse des marqueurs de la relation d'*Élaboration*.

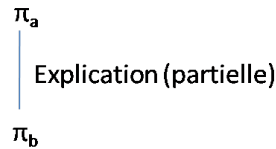


FIG. 5.3 – SDRS finale de (42)

Dans les exemples que nous allons considérer ci-dessous, *notamment* apparaît sans mention explicite du paradigme présupposé et sans marqueur lexical signalant une autre relation de discours qu'*Élaboration*.

L'adverbe *notamment* apparaît souvent avec un verbe de communication (*indiquer, préciser, dire, déclarer*) comme en (46) :

- (46) [Aux enquêteurs, elle aurait déclaré que Simone Weber avait notamment dit :]_a
 [«Ça semblerait bizarre que Bernard Hettier ne retire pas d'argent sur son compte.»]_b
 (Le Monde, 16 février 1991)

Dans ce cas, la relation qui s'établit entre le segment contenant le verbe de communication et le discours rapporté est la relation de *Source*¹ qui relie le contenu d'un acte communicatif

¹Deux nouvelles relations de discours sont apparues dans le modèle de la SDRT lors du projet d'annotation de relations de discours en corpus DISCORS (Reese *et al.*, 2007) : la relation de *Source* et la relation d'*Attribution*. Comme nous l'avons dit ci-dessus, ces deux relations sont utilisées pour relier un acte communicatif à l'agent de cet acte. La distinction entre *Source* et *Attribution* concerne les conditions de vérité du contenu du discours rapporté. Dans le cas de *Source*, l'agent s'engage sur la véracité des propos rapportés tandis que ce n'est pas le cas pour *Attribution*. *Notamment* apparaît uniquement avec des verbes engageant la relation de *Source* dans notre corpus.

à l'agent de cet acte.

Notamment spécifie que les propos rapportés ne sont qu'une partie de la totalité des propos. Toutefois la même interprétation serait obtenue sans la présence de *notamment* :

- (47) [Aux enquêteurs, elle aurait déclaré que Simone Weber avait dit :]_a [«Ça semblerait bizarre que Bernard Hettier ne retire pas d'argent sur son compte.»]_b

On ne suppose effectivement pas que tous les propos de Simone Weber ont été rapportés. Quoiqu'il en soit, on rendra compte du rôle de *notamment* au niveau du contenu propositionnel du constituant π_a en tant qu'adverbe intégré à la proposition modifieur du syntagme verbal.

Pour terminer, nous examinerons deux cas dans lesquels le segment introduit par *notamment* est attaché au contexte gauche par la relation de *Continuation* en (48) et de *Narration* en (49). En (48), le segment (d) se rattache au segment (b) par la relation de *Continuation*. Les segments (b), (c) et suivants (non mentionnés) donnent des informations sur l'évolution du joueur en équipe de France tandis que les segments (d), (e) et (f) donnent des informations sur l'évolution du joueur en club.

- (48) [Abdelkader Ben Bouali, né le 25 du mois d'Octobre de l'an 1912 à Orléansville (Algérie), était un footballeur français évoluant au poste de défenseur gauche.]_a [Il fut sélectionné une fois en équipe de France en 1937]_b [alors qu'il était sociétaire de l'Olympique de Marseille.]_c [...] [En club, il évolua notamment au Racing Universitaire d'Alger et à l'Olympique de Marseille (1935-1938).]_d [Il fut champion de France 1937 (28 matches joués sur 30 lors de la saison 37-38)]_e [et remporta la Coupe de France 1938 avec l'OM.]_f (Wikipedia)

Notamment porte sur l'éventualité «évoluer au Racing Universitaire d'Alger et à l'Olympique de Marseille». Il indique que le «Racing Universitaire d'Alger» et l'«Olympique de Marseille» ne sont pas les seuls clubs dans lesquels Abdelkader Ben Bouali a évolué durant sa carrière de footballeur.

En (49), le segment contenant *notamment* est attaché à son contexte gauche par la relation de *Narration*.

- (49) Des badauds prennent aussitôt fait et cause pour la jeune femme et se dirigent vers l'engin conduit par Djemel Bara. Quelqu'un lui aurait **notamment** expédié un coup de poing dans la figure. Bara détaille, percutant au passage la mère de la conductrice, ce qui lui occasionne des blessures induisant une interruption de travail supérieure à huit jours. (L'Est Républicain)

Notamment porte sur l'éventualité «expédier un coup de poing dans la figure». Il indique que ce «quelqu'un» a accompli d'autres actes de violence envers Djemel Bara.

Bien que cet exemple soit un exemple attesté, il nous semble qu'il serait meilleur avec la mention d'un marqueur lexical tel que *puis*.

- (50) Des badauds prennent aussitôt fait et cause pour la jeune femme et se dirigent vers l'engin conduit par Djemel Bara. **Puis**, quelqu'un lui aurait **notamment** expédié un coup de poing dans la figure. Bara détaille, percutant au passage la mère de

la conductrice, ce qui lui occasionne des blessures induisant une interruption de travail supérieure à huit jours.

Cette manipulation ne modifie pas la relation de *Narration* à l'œuvre. En revanche le connecteur *puis*, ici, en position initiale détachée, nous semble jouer le rôle d'un topique semi-implicite du type «les actions qui se sont déroulées juste après». Cela nous amènerait à considérer cet adverbe en position initiale détachée comme un introducteur de nouveau topique. Enfin, le rôle de *notamment* est d'introduire une des «actions s'étant déroulée juste après».

Ces considérations nous renvoient à l'exemple (48) pour lequel le SP «En club» peut également être traité comme un introducteur de cadre spatio-temporel (*Cf.* section 3.2.2), introduisant un nouveau topique de discours du type «les évolutions d'Abdelkader Ben Bouali lorsqu'il était en club». Le segment contenant *notamment* vient élaborer ce topique en introduisant deux des clubs dans lesquels Abdelkader Ben Bouali évolua.

Sans la mention de l'adverbial, l'emploi de *notamment* nous apparaît moins bon.

- (51) Abdelkader Ben Bouali, né le 25 du mois d'Octobre de l'an 1912 à Orléansville (Algérie), était un footballeur français évoluant au poste de défenseur gauche. Il fut sélectionné une fois en équipe de France en 1937 alors qu'il était sociétaire de l'Olympique de Marseille. Cette unique sélection date du 23 mai 1937 face à l'Irlande en match amical joué au Stade Olympique Yves-du-Manoir à Colombes (défaite 2-0 des Bleus). Ben Bouali figurait dans l'effectif de l'équipe de France à l'occasion de la phase finale de la Coupe du monde 1938, mais il ne disputa aucun des deux matchs des Bleus.

Il évolua **notamment** au Racing Universitaire d'Alger et à l'Olympique de Marseille (1935-1938). Il fut champion de France 1937 (28 matches joués sur 30 lors de la saison 37-38) et remporta la Coupe de France 1938 avec l'OM.

Nous en concluons que lorsque le segment contenant *notamment* est une poursuite du discours par *Continuation* ou par *Narration*, celui-ci sera meilleur si un segment, tel qu'un adverbial en position initiale détachée, joue le rôle de topique semi-implicite pour dominer celui-ci.

Nous allons mettre en évidence des différences entre *notamment* et les trois autres adverbes *particulièrement* (sous-section 5.2.2), *en particulier* (sous-section 5.2.3), *précisément* (sous-section 5.2.4) en s'appuyant sur les exemples avec paradigme explicite (33), (34), (35) et (39), et ceux avec paradigme implicite (42), (46) et (48).

5.2.2 *Particulièrement*

Commençons par illustrer le remplacement de *notamment* par *particulièrement* dans les exemples suivants.

- (52) ?[Des peines de six mois à huit ans de prison ferme ont été requises, mardi 29 janvier, à l'encontre des seize militants présumés du mouvement nationaliste basque Iparretarrak, jugés devant la seizième chambre du tribunal correctionnel de Paris pour «association de malfaiteurs»]_a [Le substitut du procureur de la République,

Mme Irène Stoller, a **particulièrement** réclamé une peine de huit ans d'emprisonnement contre le chef présumé du groupe Philippe Bidart, ainsi que contre Jean-Gabriel Mouesca et Joseph Etchebest.]_b

- (53) ??[En septembre 1985, un an et demi après le début de son mandat de maire, Bucaram commit une succession d'erreurs,]_a [**particulièrement** en accusant divers hommes politiques dans l'implication de meurtres mal élucidés.]_b
- (54) ??[Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [**particulièrement** à cause de scènes très crues entre Marlon Brando et Maria Schneider.]_b
- (55) ?? [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]_a [recevant **particulièrement** le représentant du gouvernement russe en Tchétchénie, Nikolaï Kochman.]_b
- (56) ?? [Aux enquêteurs, elle aurait déclaré que Simone Weber avait **particulièrement** dit :]_a [«Ça semblerait bizarre que Bernard Hettier ne retire pas d'argent sur son compte.»]_b

La substitution paraît difficile dans les 7 exemples considérés et nous remarquons que l'emploi de la forme comparative *plus particulièrement* rendrait meilleure la lecture en forçant la fonction paradigmatique de l'adverbe.

En position 3., la lecture de *particulièrement* n'est pas celle d'un adverbe paradigmatique mais plutôt celle d'un adverbe de manière quantifieur, comme l'attestent les exemples de notre corpus :

- (57) Les dispositifs de défense, mis en œuvre par ses occupants, ont **particulièrement** impressionné les élèves, qui se sont ainsi rendus compte que nos aïeux ne manquaient pas d'imagination ! (L'Est Républicain)
- (58) Les filles apprécient **particulièrement** l'espace mis à leur disposition pour installer leurs jeux. (L'Est Républicain)

En (55), on peut également voir un autre sens de *particulièrement* comme *personnellement*, *en privé*.

En revanche, la forme comparative *plus particulièrement* se substitue à *notamment* en induisant une gradation dans le paradigme présupposé. Il introduit l'élément du paradigme qui vérifie au plus haut degré la propriété définie par l'ensemble.

5.2.3 *En particulier*

Les exemples suivants montrent que *en particulier* peut se substituer à *notamment*.

- (59) [Des peines de six mois à huit ans de prison ferme ont été requises, mardi 29 janvier, à l'encontre des seize militants présumés du mouvement nationaliste basque Iparretarrak, jugés devant la seizième chambre du tribunal correctionnel de Paris pour «association de malfaiteurs»]_a [Le substitut du procureur de la République, Mme Irène Stoller, a **en particulier** réclamé une peine de huit ans d'emprisonnement contre le chef présumé du groupe Philippe Bidart, ainsi que contre Jean-Gabriel Mouesca et Joseph Etchebest.]_b

- (60) [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [**en particulier** à cause de scènes très crues entre Marlon Brando et Maria Schneider.]_b

Mais la substitution n'est pas toujours réussie comme en attestent les deux exemples suivants :

- (61) ?? [Sans arme, sans argent, ils durent se débrouiller,]_a [volant **en particulier** la voiture d'un surveillant.]_b
- (62) ?? [Aux enquêteurs, elle aurait déclaré que Simone Weber avait **en particulier** dit :]_a [«Ça semblerait bizarre que Bernard Hettier ne retire pas d'argent sur son compte.»]_b

En (63), deux interprétations concurrentes et disjointes de *en particulier* sont possibles. L'interprétation de l'adverbe paradigmatissant substituable par *entre autres* ou l'interprétation équivalente à «dans l'intimité».

- (63) [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]_a [recevant **en particulier** le représentant du gouvernement russe en Tchétchénie, Nikolai Kochman.]_b

Nous considérons les deux adverbes comme quasi-synonymes dans leurs emplois paradigmatissants. Dans ces cas, on peut également remplacer *notamment* et *en particulier* par *entre autres*.

5.2.4 Précisément

Comme nous l'avons déjà noté dans la section 5.1.4, le remplacement de *notamment* par *précisément* modifie l'interprétation, en passant de la mention d'un élément du paradigme présumé à une redescription de cet ensemble. Ce changement d'interprétation est mis en évidence dans les exemples suivants :

- (64) [En septembre 1985, un an et demi après le début de son mandat de maire, Bucaram commit une succession d'erreurs,]_a [**précisément** en accusant divers hommes politiques dans l'implication de meurtres mal élucidés.]_b
- (65) [Le Dernier Tango à Paris de Bernardo Bertolucci fait scandale,]_a [**précisément** à cause de scènes très crues entre Marlon Brando et Maria Schneider.]_b

En (66), bien que le constituant π_b ne soit pas une redescription du constituant π_a , le remplacement de *notamment* par *précisément* reste possible d'une part et ne modifie pas l'interprétation d'autre part.

- (66) [Des peines de six mois à huit ans de prison ferme ont été requises, mardi 29 janvier, à l'encontre des seize militants présumés du mouvement nationaliste basque Iparretarrak, jugés devant la seizième chambre du tribunal correctionnel de Paris pour «association de malfaiteurs»]_a [Le substitut du procureur de la République, Mme Irène Stoller, a **précisément** réclamé une peine de huit ans d'emprisonnement

contre le chef présumé du groupe Philippe Bidart, ainsi que contre Jean-Gabriel Mouesca et Joseph Etchebeste.]]_b

Cela peut s'expliquer par le fait que l'éventualité complexe décrite dans le constituant π_a («16 personnes ont reçu des peines de six mois à huit ans») et celle décrite dans π_b («3 personnes ont reçu des peines de huit ans») ne peuvent en aucun cas référer au même ensemble d'éventualités, excluant ainsi la lecture co-référentielle.

De même que pour *particulièrement*, la substitution de *précisément* à *notamment* dans les deux exemples suivants altère l'acceptabilité :

(67) ?? [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]_a [recevant **précisément** le représentant du gouvernement russe en Tchétchénie, Nikolaï Kochman.]]_b

(68) ??[Sans arme, sans argent, ils durent se débrouiller,]_a [volant **précisément** la voiture d'un surveillant.]]_b

La lecture paradigmatitante est également forcée avec la forme comparative de l'adverbe :

(69) [Le nouveau maître du Kremlin, Vladimir Poutine, a d'ailleurs consacré sa journée de lundi à s'occuper largement de ce dossier,]_a [recevant **plus précisément** le représentant du gouvernement russe en Tchétchénie, Nikolaï Kochman.]]_b

(70) [Sans arme, sans argent, ils durent se débrouiller,]_a [volant **plus précisément** la voiture d'un surveillant.]]_b

En (71), la lecture paradigmatitante de *précisément* est effacée au profit d'une lecture de *précisément* comme adverbe de manière orienté vers le sujet :

(71) [Aux enquêteurs, elle aurait déclaré que Simone Weber avait **précisément** dit :]_a [«Ça semblerait bizarre que Bernard Hettier ne retire pas d'argent sur son compte.»]]_b

Comme pour *particulièrement*, en position 3., l'emploi le plus fréquent de *précisément* n'est pas celui d'un adverbe paradigmatitant. *Précisément* est le plus souvent un adverbe de manière quantifieur, comme l'attestent les exemples de notre corpus :

(72) Revenu sur place, il a pu alors indiquer **précisément** l'endroit où il avait aperçu la carrosserie. (L'Est Républicain)

(73) Pendant des années, je ne comprenais pas ce qui se passait mais, aujourd'hui, la situation commence à se décanter. Je sais **précisément** de quoi meurent mes bêtes et je peux le prouver. (L'Est Républicain)

5.2.5 Conclusion sur les emplois intraphrastiques

Nous avons vu dans cette section 5.2 que *notamment*, *particulièrement*, *en particulier* et *précisément* ne sont pas toujours interchangeables en position 3.

Dans cette position, *particulièrement* et *précisément* sont généralement des adverbes de manière, venant modifier le syntagme verbal et ne jouant aucun rôle dans la mise en

œuvre de la relation d'*Élaboration*. Nos substitutions montrent toutefois que *précisément*, sans être paradigmatissant, peut introduire une éventualité qui est une redescription de l'éventualité précédente. Dans ce cas, la relation d'*Élaboration* est inférée.

Le rôle de *notamment* et *en particulier* en position 3. est celui d'un adverbe paradigmatissant. Nous avons dégagé deux configurations principales. Si le paradigme présupposé est explicite, on infère *Élaboration*. L'adverbe fonctionne ainsi comme un conjonctif puisque l'énoncé qu'il introduit ajoute des détails sur un énoncé précédent et son interprétation dépend de ce dernier. Si le paradigme n'est pas explicite, on n'infère jamais *Élaboration* et d'autres éléments permettent d'inférer une autre relation de discours. D'autre part, la ou les éventualités décrites dans la phrase d'accueil de *notamment* ou *en particulier* peuvent être suivies dans le discours d'une ou d'autres éventualités faisant partie du même paradigme nécessitant ainsi la construction d'un topique commun pour ces deux éventualités, Cf. (45).

Notamment, *en particulier* et *précisément* en position 3. sont des indices de la relation d'*Élaboration*.

Nous allons maintenant poursuivre l'analyse de ces adverbes en considérant leurs emplois en position initiale détachée.

5.3 Rôle discursif des adverbes en position initiale détachée

Les emplois en position initiale de nos quatre adverbes ont été peu étudiés. Ces emplois ne sont pas mentionnés dans Nolke (1983) et à peine signalés dans Molinier et Lévrier (2000). Cela peut s'expliquer par la relative récence de ces emplois en position initiale. Dans Frantext, à l'exception d'un premier emploi en position initiale de *notamment* en 1761, les emplois commencent au début du XX^{ème} siècle. Le premier emploi à l'initiale de *en particulier* est relevé en 1849, celui de *plus particulièrement* en 1907 et celui de *plus précisément* en 1902. Concernant la position initiale, nous relevons que les adverbes *particulièrement* et *précisément* apparaissent plus fréquemment dans leur forme comparative.

Dans cette section, nous allons commencer par présenter les spécificités de la position initiale. Concernant *particulièrement* et *précisément*, nous allons présenter les propriétés de leur forme comparative en opposition à leur forme simple. Enfin, nous présentons nos analyses des emplois discursifs de ces quatre adverbes.

5.3.1 Spécificité de la position initiale détachée

5.3.1.1 Emplois intraphrastiques vs. emplois à l'initiale détachée

Dans leurs emplois intraphrastiques (en position 3.), *notamment* et *en particulier* ont une fonction paradigmatissante. Le segment contenant l'adverbe peut s'attacher au contexte gauche par *Continuation* comme dans l'exemple (74).

- (74) [Comme entraîneur, il <Abdelkader Ben Bouali> est à la tête de l'équipe nationale d'Algérie qui remporte la Coupe d'Afrique des Nations en 1990.]_a [Il est **notamment** connu pour avoir refusé de sélectionner le jeune Zinedine Zidane en équipe d'Algérie, prétextant que celui-ci n'était pas assez rapide.]_b (Wikipedia)

En (74), les constituants π_a et π_b contribuent à décrire les activités de Abdelkader Ben Bouali en tant qu'entraîneur². Les deux éventualités décrites «être à la tête de l'équipe nationale d'Algérie...» et «être connu pour avoir refusé...» peuvent être interprétées comme ayant eu lieu alors qu'il était entraîneur. Les deux constituants π_a et π_b sont simplement reliés par la relation de *Continuation*.

Dans le segment (b), *notamment* signale que l'éventualité «être connu pour avoir refusé...» fait partie d'un paradigme d'éventualités plus large.

Il est impossible de déplacer *notamment* en position initiale détachée, comme le montre (75). Ce changement de position requiert un élément auquel se rattacher dans le contexte gauche, un élément qui explicite l'ensemble présupposé par *notamment* comme le montre (76) :

- (75) Comme entraîneur, il est à la tête de l'équipe nationale d'Algérie qui remporte la Coupe d'Afrique des Nations en 1990. ***Notamment**, il est connu pour avoir refusé de sélectionner le jeune Zinedine Zidane en équipe d'Algérie, prétextant que celui-ci n'était pas assez rapide.
- (76) Comme entraîneur, il est à la tête de l'équipe nationale d'Algérie qui remporte la Coupe d'Afrique des Nations en 1990. **Il a fait parler de lui à maintes reprises. Notamment**, il est connu pour avoir refusé de sélectionner le jeune Zinedine Zidane en équipe d'Algérie, prétextant que celui-ci n'était pas assez rapide.

Les adverbes *particulièrement* et *précisément*, quant à eux, se laissent beaucoup moins bien placer en position initiale contrairement à leur forme comparative, que nous allons décrire dans la section suivante.

- (77)
- | | | | |
|---|-----------------------------|-----------------------------|--------------------|
| { | Max a beaucoup d'activités. | ? Particulièrement , | il écrit un livre. |
| | | ? Précisément , | |

- (78)
- | | | | |
|---|-----------------------------|--------------------------------|--------------------|
| { | Max a beaucoup d'activités. | Plus particulièrement , | il écrit un livre. |
| | | Plus précisément , | |

5.3.1.2 Forme simple vs. forme comparative

Lorsque nous analyserons les emplois initiaux des adverbes *particulièrement* et *précisément*, nous concentrerons notre analyse sur les formes comparatives, *plus particulièrement* et *plus précisément*. La forme comparative existe seulement pour ces deux adverbes car ils sont construits sur des adjectifs scalaires, *particulier* et *précis*.

La forme comparative permet la mise en contraste explicite de deux éléments sur une échelle de valeur désignée par un terme pivot. Elle comporte trois éléments : le comparé (X), le comparant (Y) et le terme pivot (P) selon le schéma suivant :

²Dans ce cas, il nous semble pertinent de traiter le segment «Comme un entraîneur» comme un introducteur de nouveau topique de discours mais nous ne discuterons pas ici de cette question qui n'est pas notre propos dans cette section.

X verbe plus/moins/aussi P que Y

Les emplois de *plus particulièrement* et *plus précisément*, que nous nous proposons d'observer, consistent uniquement en l'expression du terme pivot. Whittaker (2004) parle dans ce cas de comparatif nu.

Il est intéressant d'étudier les formes comparatives des adverbes *particulièrement* et *précisément* pour deux raisons.

D'abord, la forme comparative peut avoir des emplois radicalement différents de la forme simple, comme le souligne Whittaker (2007, p. 19) :

- (79) A : La personne que tu décris me fait penser à Jean. B : **Précisément**, c'est de lui que je parle.
- (80) La conférence portera sur les adjectifs. **Plus précisément**, il sera question de leur emploi à la forme comparative.

Dans (79), *précisément* est un adverbe de phrase conjonctif transitionnel. On peut le remplacer par *justement*. Selon Molinier et Lévrier (2000), il a pour fonction de marquer la coïncidence entre deux situations ou deux faits ayant rapport l'un avec l'autre aux yeux du locuteur. Dans (80), *plus précisément* est la forme comparative de l'adverbe paradigmatissant.

Deuxièmement, par opposition à la forme comparative *X (le comparé) est plus précis qu'Y (le comparant)* où les contenus de X et Y sont explicites, avec le comparatif nu, le segment introduit par ce dernier correspond au comparé X tandis que la position du comparant Y est laissée vide et doit être résolue en prenant en considération le contexte en amont.

Guimier (1996) mentionne que les adverbes méta-linguistiques se présentent souvent au comparatif lorsque ceux-ci servent à la reformulation. Nous verrons si cette affirmation vaut pour les emplois de *plus particulièrement* et *plus précisément*.

Nous posons pour l'instant l'hypothèse que la combinaison de la sémantique de nos adverbes et la position initiale leur conférera un rôle de connexion. L'hypothèse est encore plus forte pour les formes comparatives, étant donné que la forme du comparatif nu laisse la position du comparé vide, qui doit alors être remplie anaphoriquement par un élément du contexte gauche.

5.3.2 *Notamment, en particulier*

Nous analysons de façon groupée les adverbes *notamment* et *en particulier* dont nous avons montré les similitudes d'emploi lorsqu'ils se trouvent en position 3.

- (81) [Cette dureté intervient aussi dans la précision que l'on peut atteindre avec des outils en diamant.]_a [**Notamment**, les scalpels en diamant permettent de créer des incisions ultraprécises (en ophtalmologie par exemple)]_b. (Wikipedia)

- (82) [Le plumage du cormoran aptère est noir avec des reflets bruns et quelques plumes blanchâtres.]_a [Son anatomie est différente de la plupart des oiseaux.]_b [**En particulier**, son sternum n'a plus de bréchet sur lequel viennent normalement s'accrocher les muscles des ailes.]_c (Wikipedia)

En (81), le prédicat *Subtype_D* est vérifié entre «la précision que l'on peut atteindre avec des outils en diamant» et «le scalpel en diamant permet de créer des incisions ultraprécises» sur la base de connaissances lexicales selon lesquelles «un scalpel en diamant» est un type d'outil. Cela permet d'inférer la relation d'*Élaboration* entre π_a et π_b sans avoir recours à *notamment* dont la suppression est parfaitement possible :

- (83) [Cette dureté intervient aussi dans la précision que l'on peut atteindre avec des outils en diamant.]_a [Les scalpels en diamant permettent de créer des incisions ultraprécises (en ophtalmologie par exemple)]_b.

L'emploi de *notamment* en position initiale détachée dans le segment (b) fait intervenir d'une part un paradigme d'éventualités (du type «des outils en diamants permettent de réaliser des tâches de précision») et d'autre part la nécessité que ce paradigme soit précédemment explicité dans le contexte gauche.

Dans (82), le prédicat *Subtype_D* est vérifié entre «l'anatomie du cormoran est différente de celle de la plupart des oiseaux» et «son sternum n'a plus de bréchet sur lequel viennent normalement s'accrocher les muscles des ailes» sur la base de connaissances du monde selon lesquels la plupart des oiseaux ont un bréchet sur lequel les muscles des ailes viennent s'accrocher. Ce qui n'est pas le cas chez le cormoran et constitue donc une différence de ce dernier vis à vis de la plupart des oiseaux. Cela permet d'inférer *Élaboration*(π_b, π_c). Cela étant dit, même sans ces connaissances préalables, le locuteur peut inférer la relation d'*Élaboration*, i.e. inférer que «le sternum n'a plus de bréchet» différencie le cormoran des autres animaux. Le segment (a) indique qu'il existe des différences anatomiques entre le cormoran et les autres oiseaux et le segment (b) indique une caractéristique anatomique. Le locuteur infère, avec et sans la présence de *en particulier*, que le segment (b) indique une de ces différences anatomiques.

- (84) [Son anatomie est différente de la plupart des oiseaux.]_b [Son sternum n'a plus de bréchet sur lequel viennent normalement s'accrocher les muscles des ailes.]_c

Comme nous le voyons avec les exemples (81) et (82), même en l'absence de connaissances spécifiques, le lecteur peut fournir les inférences nécessaires afin de déterminer un élément du contexte gauche qui peut jouer le rôle de topique discursif. La cohérence constitue un principe premier d'interprétation des discours et l'emploi de *notamment* et *en particulier* en position initiale détachée nécessite qu'un topique discursif soit précédemment explicitement cité. Si l'on prend un cas extrême comme ci-dessous :

- (85) Max a vécu plusieurs catastrophes l'an dernier.

Notamment,
En particulier, il s'est marié en août.

Bien que «se marier» ne soit ni lexicalement, ni dans les connaissances du monde, un événement susceptible d'être une catastrophe, la présence de *notamment* ou *en particulier* force, en quelque sorte, cette interprétation. Le lecteur va imaginer par exemple qu'un événement qui eut lieu en août tel qu'une catastrophe naturelle a ruiné le mariage ou plus simplement que le mariage a été une catastrophe. En l'absence de l'un des deux adverbes (Cf. exemple (86)), cette interprétation demeure possible mais l'inférence la plus naturelle semble d'interpréter une *Continuation* entre π_a et π_b , comme si le scripteur faisait une liste des événements marquants de la vie de Max :

(86) [Max a vécu plusieurs catastrophes l'an dernier.]_a [Il s'est marié en août.]_b

La présence de *notamment* ou *en particulier* force l'interprétation du deuxième segment comme une élaboration du premier et bloque l'interprétation de continuation.

Bilan

Tous ces exemples et ces manipulations nous amènent à considérer un rôle fort des adverbes *notamment* et *en particulier* en position initiale détachée. Leur présence en tête de phrase nécessite que soit mentionné dans le contexte gauche un élément jouant le rôle de topique discursif et qui se trouve élaboré par le segment introduit par l'un des deux adverbes. *Notamment* et *en particulier* en position initiale détachée sont des marqueurs spécificateurs de la relation d'*Élaboration*.

Axiome 5.2 Inférer ÉlaborationNotamment PID

$(?(\alpha, \beta, \lambda) \wedge [\text{Notamment},](\beta)) \rightarrow \text{Élaboration}(\alpha, \beta)$

Axiome 5.3 Inférer Élaboration En particulier PID

$(?(\alpha, \beta, \lambda) \wedge [\text{En particulier},](\beta)) \rightarrow \text{Élaboration}(\alpha, \beta)$

Du point de vue de l'interprétation et d'une analyse SDRT, *notamment* et *en particulier*, bien que généralement non nécessaires, facilitent l'interprétation dans la mise en œuvre de la relation d'*Élaboration*. Du point de vue du repérage automatique, ils sont considérés comme des indices fiables de celle-ci.

5.3.3 Plus particulièrement

Cette section va concerner l'adverbe *particulièrement* au comparatif nu. Cette forme est propice au rôle de connexion étant donné que la position du comparé est laissée vide et résolue anaphoriquement par la présence d'un élément dans le contexte gauche.

(87) [De Halve Maen est un périodique historique étatsunien (publié à New York) par la Holland Society of New York depuis 1922.]_a [Publication trimestrielle, différents universitaires et historiens y font connaître le fruit de leurs recherches concernant l'action coloniale néerlandaise du XVIIe siècle.]_b [**Plus particulièrement**, l'histoire des Nouveaux-Pays-Bas y est décortiquée.]_c

En (87), le terme pivot «plus particulièrement» définit la relation de comparaison entre le comparé, qui correspond à la phrase introduite par le terme pivot, «l’histoire des Nouveaux-Pays-Bas y est décortiquée», et le comparant, qui correspond à la phrase précédente, «différents universitaires et historiens y font connaître le fruit de leurs recherches concernant l’action coloniale néerlandaise du XVII^e siècle».

Une analyse SDRT de cet exemple exploiterait des connaissances du monde (que les Nouveaux-Pays-Bas sont une ancienne colonie néerlandaise) pour inférer *Subtype_D* entre le type de l’éventualité décrite dans π_c «décortiquer l’histoire des Nouveaux-Pays-Bas» et dans π_b «faire connaître les recherches entre l’action coloniale néerlandaise», puis inférer *Élaboration*(π_b, π_c).

Concernant maintenant le rôle de *plus particulièrement*, on reconnaît son rôle paradigmatissant. Il signifie que des recherches sont menées sur les Nouveaux-Pays-Bas à un plus haut degré que les autres actions coloniales néerlandaises. L’adverbe peut être déplacé à l’intérieur de la phrase sans en affecter le sens :

- (88) Publication trimestrielle, différents universitaires et historiens y font connaître le fruit de leurs recherches concernant l’action coloniale néerlandaise du XVII^e siècle. L’histoire des Nouveaux-Pays-Bas y est **plus particulièrement** décortiquée.

Nous allons poursuivre l’analyse de *plus particulièrement* en distinguant deux types d’emploi de l’adverbe, d’un côté des emplois paradigmatissants et conjonctifs et de l’autre des emplois où la fonction paradigmatissante semble disparaître.

5.3.3.1 *plus particulièrement* paradigmatissant et conjonctif

Dans les exemples suivants, *plus particulièrement* a la même fonction que dans (87).

- (89) [À cause de sa silhouette reconnaissable et visible de loin, la Cathédrale Saint-Paul sert souvent à situer l’action à Londres (Comme la tour Eiffel à Paris).]_a
[Plus particulièrement, la cathédrale apparaît dans :
 - Mary Poppins (La chanson *Feed the Birds* y a été tournée)
 - Steamboy
 - Les 101 dalmatiens
 - Peter Pan]_b

Il a une fonction paradigmatissante, signalant que la «Cathédrale Saint-Paul» sert souvent à situer l’action à Londres dans les films cités mais également dans d’autres films, de la même façon que si l’adverbe occupe une position intraphrastique (au moins pour la fonction paradigmatissante) :

- (90) La cathédrale apparaît **plus particulièrement** dans :

Nous notons que le segment (b) est un macro-segment correspondant à une structure énumérative (cf. chapitre 6 sur les structures énumératives), constituée d’une amorce («Plus particulièrement, la cathédrale apparaît dans») et d’une énumération (liste d’items introduits par des tirets). *Plus particulièrement* permet de spécifier que l’énumération des films n’est qu’un sous-ensemble des films dans lesquels apparaît la Cathédrale Saint-Paul.

Observons un deuxième exemple.

- (91) [Le Guide Michelin est le nom générique d'une série de guides publiée annuellement par Michelin pour une douzaine de pays différents.]_a [**Plus particulièrement**, l'expression désigne le guide rouge de Michelin, le plus ancien et le plus fameux des guides gastronomiques européens.]_b

Plus particulièrement a une fonction paradigmatitante, en faisant la sélection d'un guide Michelin (le guide rouge) parmi l'ensemble des guides. Cet énoncé antérieur se manifeste dans le segment (a) dont l'éventualité décrit «le Guide Michelin comme le nom générique d'une série de guides». Du point de vue d'une analyse SDRT, la relation inférée entre π_a et π_b est *Élaboration*.

Dans les cas que nous venons d'étudier, nous pouvons substituer *notamment* ou *en particulier* à *plus particulièrement* :

- (92) [À cause de sa silhouette reconnaissable et visible de loin, la Cathédrale Saint-Paul sert souvent à situer l'action à Londres (Comme la tour Eiffel à Paris).]_a
[**Notamment/En particulier**, la cathédrale apparaît dans :
- Mary Poppins (La chanson Feed the Birds y a été tournée)
- Steamboy
- Les 101 dalmatiens
- Peter Pan]_b
- (93) [Le Guide Michelin est le nom générique d'une série de guides publiée annuellement par Michelin pour une douzaine de pays différents.]_a [**Notamment/En particulier**, l'expression désigne le guide rouge de Michelin, le plus ancien et le plus fameux des guides gastronomiques européens.]_b

Contrairement aux emplois intraphrastiques de *particulièrement* pour lesquels Molinier et Lévrier (2000) indiquaient que «l'ensemble doit être construit sur une propriété gradable et le sous-ensemble se distingue dans l'ensemble de référence par le degré auquel ses éléments vérifient cette propriété.», nous notons que l'éventualité introduite par *plus particulièrement* ne se distingue pas de l'ensemble présupposé par le degré auquel cette ou ces éventualités vérifient la propriété définie par l'ensemble présupposé. En clair, lors de l'interprétation de (89), l'ensemble «la cathédrale apparaît dans les films pour situer l'action à Londres» n'est pas construit sur une propriété gradable et cette propriété ne se vérifie pas plus pour les films cités que pour les autres films du paradigme. De même pour (91), l'ensemble «le Guide Michelin est le nom générique d'une série de guides» n'est pas construit sur une propriété gradable et cette propriété ne se vérifie pas plus pour le guide rouge que pour les autres guides de la gamme Michelin.

La perte de cette propriété gradable peut s'expliquer par la spécificité du comparatif nu qui met en relation, non pas les éléments de l'ensemble (comme *particulièrement* intraphrastique impose une échelle gradable entre les éléments de l'ensemble présupposé), mais un comparé (l'ensemble présupposé) et un comparant (le sous-ensemble désigné par *plus particulièrement*) selon une relation établie par le terme pivot («plus particulièrement») : l'information fournie dans le comparé est plus *particulière* que l'information fournie dans le

comparant. *Particulier* est ici pris dans le sens contraire de *général* : l'information fournie dans le comparé est plus *générale* que l'information fournie dans le comparant. On voit ainsi s'opposer les termes de *souvent* et *plus particulièrement* en (89) et les termes *générique* et *plus particulièrement* en (91).

Dans les premiers exemples (87), (89) et (91) que nous avons analysés, le terme pivot *plus particulièrement* met en relation l'ensemble présupposé (le comparant) et un élément ou sous-ensemble (le comparé). Deux constituants sont ainsi reliés par la relation d'*Élaboration* et leur contenu se caractérise par une mise en relation du général au particulier. Dans les prochains exemples que nous allons observer, le terme pivot *plus particulièrement* permet de mettre en relation deux éléments et/ou sous-ensembles de l'ensemble présupposé par l'adverbe, ce qui correspond intuitivement au sens de l'adverbe *particulièrement* (établir une hiérarchie entre des éléments d'un ensemble).

- (94) [Alors que les signaux numériques sont en général associés avec les systèmes électroniques numériques binaires utilisés dans l'électronique moderne et dans l'informatique,]_a
 [bien des systèmes numériques ont existé pendant les siècles précédents et n'étaient en aucun cas électroniques.]_b
 [Une balise lumineuse est peut-être l'exemple le plus simple d'un signal numérique qui n'est pas électronique, avec deux états (marche et arrêt).]_c [**Plus particulièrement**, les signaux de fumée sont l'un des exemples les plus vieux des signaux numériques, où un porteur analogique (la fumée) est modulé par une couverture afin de générer un signal numérique qui transporte les informations.]_d

Dans l'exemple (94), le comparatif nu *plus particulièrement* met en relation un comparé «la balise lumineuse» et un comparant «les signaux de fumée» : les deux sont des sous-types de «systèmes numériques non électroniques» (du segment b). L'emploi de *plus particulièrement* est ici métadiscursif : le scripteur accorde un intérêt particulier aux signaux de fumée par rapport aux balises lumineuses en tant qu'exemplaire de la classe des signaux numériques non électroniques.

Il s'ensuit que le rôle discursif de *plus particulièrement* n'est pas le même que celui que nous avons décrit précédemment. Le segment auquel il se rattache dans le contexte gauche décrit un autre élément du paradigme présupposé par l'adverbe.

Dans le cadre d'une analyse SDRT, le constituant π_d introduit par *plus particulièrement* est attaché au constituant π_c par la relation de *Parallèle* (inférée principalement sur l'isomorphisme syntaxique des deux structures «être l'exemple le plus...»). Le constituant π_c est rattaché au constituant π_b par la relation d'*Élaboration* (le segment (b) spécifie que des signaux numériques non électroniques ont existé et le segment (c) introduit un exemple de ces signaux numériques). Il s'ensuit, par le principe inverse de distributivité que π_d attaché à une relation coordonnante à π_b poursuit la même fonction rhétorique que celui-ci. On infère donc *Élaboration*(π_b, π_d).

En (95), *plus particulièrement* présente le même fonctionnement :

- (95) [En adressant cette réflexion à nos frères et sœurs de l'Eglise catholique à travers le monde,]_a [nous invitons tous les chrétiens à nous rejoindre et à méditer sur la

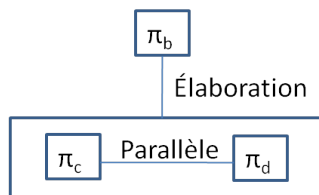


FIG. 5.4 – Représentation discursive de (94)

catastrophe qui est advenue au peuple juif et à l'impératif moral pour s'assurer que, jamais plus, l'égoïsme et la haine ne prospèrent jusqu'au point de semer tant de souffrance et de mort.]]_b [**Plus particulièrement**, nous invitons nos amis juifs dont «le destin terrible est devenu un symbole des aberrations dont l'homme est capable quand il se retourne contre Dieu» (Jean Paul II, allocution aux chefs de la communauté juive de Budapest, 18 août 1991) à nous écouter avec générosité de cœur.]]_c (Wikipedia)

Le comparatif nu *plus particulièrement* met en relation un comparé «tous les chrétiens» et un comparant «nos amis juifs». Les deux forment l'ensemble présupposé par *plus particulièrement* des personnes invitées par le Pape. L'emploi de *plus particulièrement* indique que les «amis juifs» sont *plus particulièrement* que les «chrétiens» invités à écouter l'allocution du Pape. Le segment auquel se rattache (c) est le segment (b) qui mentionne un autre sous-ensemble de l'ensemble présupposé par *plus particulièrement*.

Comme pour l'analyse de (94), le constituant π_d introduit par *plus particulièrement* est attaché au constituant π_c par la relation de *Parallèle* (inférée principalement sur l'isomorphisme syntaxique des deux structures «nous invitons... à...»).

Dans ces configurations là, *plus particulièrement* ne peut évidemment pas être remplacé par *notamment* ou *en particulier* :

- (96) [Alors que les signaux numériques sont en général associés avec les systèmes électroniques numériques binaires utilisés dans l'électronique moderne et dans l'informatique,]]_a
 [bien des systèmes numériques ont existé pendant les siècles précédents et n'étaient en aucun cas électroniques.]]_b
 [Une balise lumineuse est peut-être l'exemple le plus simple d'un signal numérique qui n'est pas électronique, avec deux états (marche et arrêt).]]_c
 [***Notamment/*En particulier**, les signaux de fumée sont l'un des exemples les plus vieux des signaux numériques, où un porteur analogique (la fumée) est modulée par une couverture afin de générer un signal numérique qui transporte les informations.]]_d
- (97) [En adressant cette réflexion à nos frères et sœurs de l'Eglise catholique à travers le monde,]]_a [nous invitons tous les chrétiens à nous rejoindre et à méditer sur la catastrophe qui est advenue au peuple juif et à l'impératif moral pour s'assurer que, jamais plus, l'égoïsme et la haine ne prospèrent jusqu'au point de semer tant de souffrance et de mort.]]_b [***Notamment/*En particulier**, nous invitons nos amis juifs dont «le destin terrible est devenu un symbole des aberrations dont l'homme est capable quand il se retourne contre Dieu» (Jean Paul II, allocution

aux chefs de la communauté juive de Budapest, 18 août 1991) à nous écouter avec générosité de cœur.]_c (Wikipedia, traduit de l'italien)

Bilan

Nous avons mis en évidence deux rôles de l'adverbe paradigmatissant *plus particulièrement*.

Dans le premier cas de figure illustré en (89) et (91), le rattachement du segment d'accueil de *plus particulièrement* se fait *via* la relation de discours *Élaboration*. C'est à chaque fois le cas quand la position vide du comparant est remplie par un constituant décrivant l'ensemble présupposé, Cf. Figure 5.5.

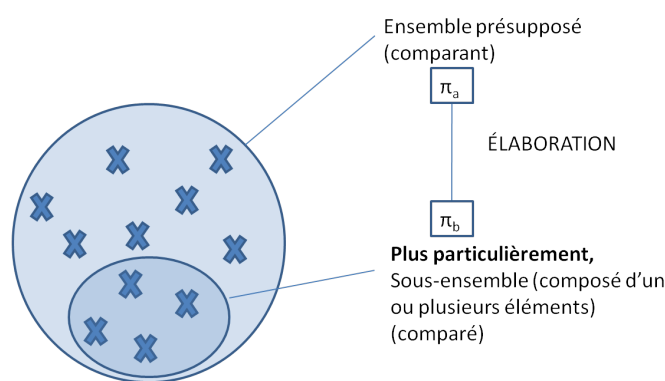


FIG. 5.5 – Rôle paradigmatissant et conjonctif de *plus particulièrement* (1)

Dans le second cas de figure illustré en (94) et (95), le rattachement du segment d'accueil de *plus particulièrement* se fait *via* une relation coordonnante³. C'est à chaque fois le cas quand la position vide du comparant est remplie par un constituant décrivant un élément ou sous-ensemble de l'ensemble présupposé par *plus particulièrement*, Cf. Figure 5.6.

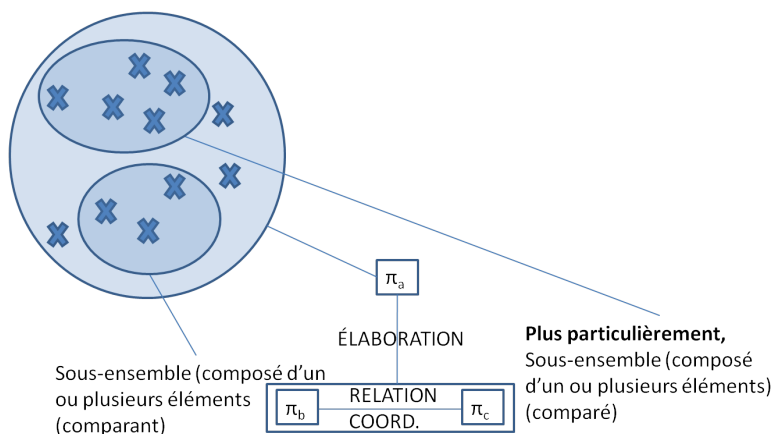
Dans ce second cas, *plus particulièrement* a également en position initiale détachée une fonction paradigmatissante mais l'instruction de rattachement n'est pas exactement la même. *plus particulièrement* relie deux éléments de l'ensemble présupposé, qui se trouvent alors au même niveau discursif et peuvent être dominés par un segment désignant cet ensemble.

Dans la prochaine section, nous allons étudier des emplois de *plus particulièrement* conjonctifs mais dont la fonction paradigmatissante semble s'effacer.

5.3.3.2 Autres emplois de *plus particulièrement*

Nous avons montré à plusieurs reprises le rôle conjonctif de *plus particulièrement* en position initiale détachée. En tant qu'élément signalant la connexion entre deux éléments

³Nous n'avons pas suffisamment d'exemples dans notre corpus pour conclure que le rattachement se fait toujours par *Parallèle*. Selon nous, prime l'idée que les deux constituants décrivent des éventualités qui se situent au même niveau d'analyse.

FIG. 5.6 – Rôle paradigmatissant et conjonctif de *plus particulièrement* (2)

textuels, il a le rôle de spécifier la nature du rapport entre les énoncés connectés. Dans les cas canoniques, *plus particulièrement* spécifie, dans un ensemble d'éventualités, une ou plusieurs éventualités comme nous l'avons vu dans les analyses des sections précédentes. Mais dans certains emplois que nous étudions dans cette section, la fonction paradigmatissante semble progressivement s'effacer.

- (98) [D'une manière générale le mot flexion désigne l'action de fléchir, de courber, de plier ou son résultat.]_a [**Plus particulièrement**, il est employé dans les domaines suivants :
- en science des matériaux, la flexion est un mode de déformation de la matière ;
 - en linguistique, la flexion est la variation de la forme des mots selon le contexte (conjugaison, déclinaison) ;
 - en anatomie, une flexion est une action consistant à plier un membre par opposition à l'extension.]_b (Wikipedia)

En (98), le rôle de *plus particulièrement* n'est plus de spécifier une éventualité ou un sous-ensemble d'éventualités dans un paradigme d'éventualité, explicité dans le segment précédent mais uniquement de connecter deux segments, laissant penser que l'adverbe évolue vers un rôle purement conjonctif.

De nombreux cas reproduisent le schéma de l'exemple (98). *Plus particulièrement* est le pendant d'une expression introduisant un segment d'information générale, comme *d'une manière générale* dans (98) ou *en règle générale*. Il permet, ici, d'introduire les définitions du terme «flexion» dans des domaines de spécialité (par exemple les sciences des matériaux, la linguistique, l'anatomie...) par opposition à la définition générale du terme.

Plus particulièrement a un emploi paradigmatissant. Mais, dans sa fonction conjonctive, il ne permet plus de connecter un sous-ensemble d'éventualités extrait de l'ensemble présupposé (i.e. «tous les domaines dans lesquels «flexion» a un sens particulier») à ce paradigme explicité dans le contexte gauche. La conjonction est opérée en opposant l'emploi général de «flexion» et les emplois dans des domaines de spécialité qui découlent de cet emploi général.

Dans ce cas, la substitution avec *notamment* est exclue mais possible avec *en particulier* (qui peut également fonctionner en tandem avec des expressions tels *d'une manière générale, en général, en règle générale...*) :

- (99) [D'une manière générale le mot flexion désigne l'action de fléchir, de courber, de plier ou son résultat.]_a [***Notamment, /En particulier**, il est employé dans les domaines suivants :
- en science des matériaux, la flexion est un mode de déformation de la matière ;
 - en linguistique, la flexion est la variation de la forme des mots selon le contexte (conjugaison, déclinaison) ;
 - en anatomie, une flexion est une action consistant à plier un membre par opposition à l'extension.]_b

En SDRT, le prédicat *Subtype_D* n'est pas vérifié entre π_a et π_b , rendant impossible l'inférence de la relation d'*Élaboration* en l'absence de prise en compte de l'adverbe. Dans le cadre de la RST, les deux segments sont reliés par *Élaboration* selon le sous-type appelé *General-Specific*. La question de la relation qui unit ces deux segments se pose réellement dans le cadre de la SDRT. La relation entre une information ou des informations spécifiques à une information générale pourrait, nous semble-t-il, être traitée sous l'étendard de l'*Élaboration*.

Dans l'exemple (100), *plus particulièrement* ne permet pas de sélectionner une éventualité ou un ensemble d'éventualités de l'ensemble présupposé par *plus particulièrement*, il permet au contraire d'introduire une redescription.

- (100) [GEM]_a
 [Grenoble Ecole de Management est l'institution «mère», regroupant l'ensemble des programmes de formation nationaux et internationaux du groupe et ses activités transversales : relation entreprises, entrepreneuriat, recherche appliquée, etc.]_b [**Plus particulièrement**, GEM est constitué des entités suivantes :
- ESC Grenoble ;
 - Grenoble Graduate School of Business ;
 - EMSI, École de Management des Systèmes d'Information]_c (Wikipedia)

Le verbe «être constitué de» indique que les entités énumérées correspondent à l'ensemble des entités composantes de GEM.

Le segment (c) («GEM est constitué de ESC Grenoble, Grenoble Graduate School of Business et EMSI») apporte avec plus de détails les informations contenues dans le segment (b) («GEM regroupe l'ensemble des programmes de formation nationaux et internationaux»). Il est possible de remplacer *plus particulièrement* par *plus précisément* mais pas par *notamment* :

- (101) [GEM]_a
 [Grenoble Ecole de Management est l'institution «mère», regroupant l'ensemble des programmes de formation nationaux et internationaux du groupe et ses activités transversales : relation entreprises, entrepreneuriat, recherche appliquée, etc.]_b [**Plus précisément, /*Notamment**, GEM est constitué des entités suivantes :

- ESC Grenoble ;
- Grenoble Graduate School of Business ;
- EMSI, École de Management des Systèmes d'Information]_c

Dans ce cas, *plus particulièrement* semble avoir totalement perdu sa fonction paradigmatitante. Il en résulte, de notre point de vue, une sorte de contradiction, nous amenant à penser qu'il peut s'agir d'un emploi déviant.

Dans le cadre d'une analyse SDRT, il est possible d'inférer *Subtype_D* entre le type de l'éventualité décrite en (c) «GEM est constitué de ESC Grenoble, Grenoble Graduate School of Business et EMSI») et le type de l'éventualité décrite en (b) «GEM regroupe l'ensemble des programmes de formation nationaux et internationaux». On peut donc inférer *Élaboration*(π_b, π_c) sans tenir compte de *plus particulièrement* dont la suppression améliore, de notre point de vue, la qualité de cet extrait.

L'interprétation de *plus particulièrement* présente des caractéristiques stables permettant de conclure qu'il est un marqueur spécifique de la relation d'*Élaboration* ou de la relation de *Continuation*, moyennant la prise en considération du contexte dans lequel il est utilisé. Ces différentes interprétations ont été mises en évidence avec la possibilité ou l'impossibilité de remplacer *plus particulièrement* avec les autres adverbes étudiés dont *plus précisément* que nous allons considérer dans la section suivante.

5.3.4 *Plus précisément*

Comme *plus particulièrement*, *plus précisément* établit une relation de comparaison entre un comparé (introduit dans la phrase d'accueil) et un comparant (défini dans le contexte gauche), ce que nous pouvons observer en (102) et (103).

- (102) [En matière religieuse, il <Innocent VIII> publie la bulle Summis desiderantes affectibus (5 décembre 1484), autorisant l'Inquisition à agir en matière de sorcellerie.]_a
[Plus précisément], elle autorise Henri Institoris et Jacques Sprenger à instruire le procès de deux sorcières présumées en Allemagne.]_b
- (103) [M. Jean Weber, inspecteur des finances et PDG du Crédit Industriel d'Alsace, a été chargé par le premier ministre, M. Michel Rocard, d'une mission d'expertise dans le domaine du médicament.]_a
[Plus précisément], il est chargé de préparer les statuts de la future agence du médicament et d'élaborer le nouveau système de fixation des prix des produits pharmaceutiques, dont l'expérimentation sera faite durant le deuxième trimestre de 1991.]_b (Le Monde)

En (102), le terme pivot est *plus précisément*, il définit la relation de comparaison entre le comparant, «la bulle autorise Henri Institoris et Jacques Sprenger à instruire le procès de deux sorcières présumées en Allemagne», et le comparé «la bulle autorise l'Inquisition à agir en matière de sorcellerie».

Une analyse SDRT de cet exemple exploiterait des connaissances du monde pour inférer *Subtype_D* entre «la bulle autorise Henri Institoris et Jacques Sprenger à instruire le procès de deux sorcières présumées en Allemagne» et «la bulle autorise l'Inquisition à agir en matière de sorcellerie», puis pour inférer *Élaboration*(π_a, π_b).

Le rôle de *plus précisément* est de spécifier que le contenu du comparant est plus *précis* que le contenu du comparé en ce qui concerne les instructions de la Bulle. L'adverbe peut être déplacé à l'intérieur de la phrase sans en affecter le sens :

- (104) En matière religieuse, il <Innocent VIII> publie la bulle *Summis desiderantes affectibus* (5 décembre 1484), autorisant l'Inquisition à agir en matière de sorcellerie. Elle autorise **plus précisément** Henri Institoris et Jacques Sprenger à instruire le procès de deux sorcières présumées en Allemagne.

En (103), l'information fournie dans le comparé («être chargé de préparer les statuts de la future agence du médicament et d'élaborer le nouveau système de fixation des prix») est plus *précise* que l'information fournie dans le comparant («être chargé d'une mission d'expertise dans le domaine du médicament»).

Le prédicat *Subtype_D* est inféré entre les deux éventualités «être chargé de préparer les statuts de la future agence du médicament et d'élaborer le nouveau système de fixation des prix» et «être chargé d'une mission d'expertise dans le domaine du médicament», «préparer les statuts» et «élaborer un système financier» étant dans les connaissances du monde des sous-types de «mission». La relation inférée est donc *Élaboration*(π_a, π_b). L'inférence de la relation est basée sur les types des éventualités reliées qui sont des sous-types, sans prendre en compte le rôle de *plus précisément* dont on remarque qu'il peut être supprimé :

- (105) [M. Jean Weber, inspecteur des finances et PDG du Crédit Industriel d'Alsace, a été chargé par le premier ministre, M. Michel Rocard, d'une mission d'expertise dans le domaine du médicament.]_a
[Il est chargé de préparer les statuts de la future agence du médicament et d'élaborer le nouveau système de fixation des prix des produits pharmaceutiques, dont l'expérimentation sera faite durant le deuxième trimestre de 1991.]_b

Toutefois, *plus précisément* joue le rôle de spécificateur de la relation, en explicitant la relation d'*Élaboration* à travers l'instruction donnée par le comparatif nu : Y est plus précis que X.

Examinons un autre exemple d'emploi de *plus précisément*, (106), qui se laisse analyser différemment en SDRT.

- (106) [Pour certaines oeuvres la forme matérielle du livre, telle qu'elle a été prévue ou rêvée par l'écrivain, importe presque autant que son contenu.]_a [**Plus précisément**, elle constitue l'accompagnement nécessaire de l'œuvre, son écrin adéquat.]_b (Le Monde)

Ici, *Subtype_D* ne peut pas être directement inféré entre «la forme matérielle constitue l'accompagnement nécessaire de l'œuvre» et «la forme matérielle du livre importe autant que son contenu». «Être l'accompagnement de l'œuvre» n'est pas un sous-type d'«importer autant que l'œuvre».

De notre point de vue, *plus précisément*, apportant l'instruction «être l'accompagnement de l'œuvre» est une information plus précise qu'«importer autant que l'œuvre», déclenche l'inférence d'*Élaboration*. Plus exactement, nous avons, ici, une forme de correction du premier segment avec le contenu du second.

Plus précisément agit donc comme apportant dans le discours un élément de précision et/ou de correction vis à vis de l'éventualité précédemment introduite, ce que montrent plus avant les exemples (107) et (108) :

- (107) [Il s'appelait Le Secq.]_a [**Plus précisément**, Henri Le Secq des Tournelles.]_b (Le Monde)
- (108) [Bizarrement, les programmeurs ont choisi d'inaugurer cette case avec une rediffusion.] [Anges et démons de la cité, le remarquable document réalisé par Frédéric Laffont, déjà diffusé sur France 3 en 1994 dans le magazine «Du côté de Zanzi bar» (disparu de la grille la même année), ouvre la série.]_a [**Plus précisément**, la chaîne qui omet de le signaler, propose une version écourtée du documentaire original, réduit de trente minutes pour la circonstance.]_b (Le Monde)

En (107) et (108), deux éventualités sont décrites, dans les segments (a) et (b), seulement l'information contenue dans le segment (b) est plus précise et est introduite comme venant compléter les informations précédemment introduites. *Plus précisément* joue un rôle au niveau de l'énonciation. Parmi les adverbes à portée sur le dire (Guimier, 1996), il se range dans les adverbes métalinguistiques. Le rôle de ces adverbes est de permettre à l'énonciateur de donner des indications relatives à la forme linguistique. L'instruction donnée par *plus précisément* revient à «ce que je te dis en Y est plus précis que ce que je te dis en X». Dans ce cas, la relation de discours s'établissant entre X et Y se distingue du cas classique d'élaboration entre événements. Il s'agit plutôt d'une élaboration s'établissant au niveau énonciatif, comme une sorte de méta-élaboration.

L'éventualité introduite par *plus précisément* est toujours une redescription (avec plus de détails, plus d'informations) de l'éventualité précédemment introduite. Cette observation permet maintenant d'expliquer pourquoi *plus précisément* ne peut pas se substituer à *plus particulièrement* dans l'exemple (95) :

- (109) (...) nous invitons tous les chrétiens à nous rejoindre et à méditer sur la catastrophe (...) ***Plus précisément**, nous invitons nos amis juifs (...) à nous écouter avec générosité de cœur

Nous venons de montrer que *plus précisément* peut mettre en relation :

- deux descriptions de deux éventualités dont l'une est décrite avec plus de précision, ce que ne peuvent pas faire *notamment* et *en particulier* :

(110)

Il s'appelait Secq. ***Notamment**, Henri le Secq des Tournelles.
***En particulier**,

- deux descriptions de deux éventualités dont la première est un type d'éventualité divisible («être chargé d'une mission» peut se subdiviser en plusieurs actions). Dans ce cas la substitution avec *notamment* et *en particulier* est possible :

- (111) [M. Jean Weber, inspecteur des finances et PDG du Crédit Industriel d'Alsace, a été chargé par le premier ministre, M. Michel Rocard, d'une mission d'expertise dans le domaine du médicament.]_a

[**Notamment/En particulier**, il est chargé de préparer les statuts de la future agence du médicament et d'élaborer le nouveau système de fixation des prix des produits pharmaceutiques, dont l'expérimentation sera faite durant le deuxième trimestre de 1991.]_b

Toutefois, avec *notamment* et *en particulier*, l'éventualité introduite par l'adverbe constitue un sous-ensemble du paradigme déterminé dans le contexte gauche «être chargé d'une mission d'expertise».

La relation de discours s'établit toujours selon la relation de comparaison instruite par le comparatif nu *plus précisément* : Y (le comparant) est plus précis que X (le comparé). Nous postulons donc pour un rôle fort de *plus précisément* dans la mise en œuvre de la relation d'*Élaboration* bien que ce type d'élaboration diffère des cas classiques. Celui-ci est un marqueur spécificateur de la relation d'*Élaboration*.

Axiome 5.4 Inférer Élaboration Plus précisément PID
 $(?(\alpha, \beta, \lambda) \wedge [Plus\ précisément,](\beta)) > \acute{E}laboration(\alpha, \beta)$

5.3.5 Conclusion sur les emplois à l'initiale

Parmi les adverbes paradigmatiques que nous avons analysés dans ce chapitre, nous avons distingué des différences dans les fonctionnements discursifs.

- *Notamment* isole un élément comme exemplaire d'un ensemble, Cf. l'axiome 5.2 page 154 pour inférer *Élaboration*.

- *En particulier* peut avoir deux fonctions : soit il isole un élément comme exemplaire d'un ensemble gradué ; soit il oppose une information générale à une information particulière, Cf. l'axiome 5.3 page 154.

- *Plus particulièrement* isole un élément comme exemplaire d'un ensemble gradué : soit il met cet élément en relation avec l'ensemble et on infère *Élaboration* ; soit il met cet élément en relation avec un autre élément de l'ensemble et on infère *Continuation*. Il peut également, comme *en particulier*, opposer le général au particulier.

- *Plus précisément* peut avoir deux fonctions : soit il isole une information comme étant plus précise que l'information précédemment donnée. Dans ce cas, il peut être remplacé par *notamment* et *en particulier* bien que cela entraîne un changement de sens. Dans ce cas on infère *Élaboration*. Soit il fonctionne comme un adverbe métalinguistique. Il introduit de cette façon une correction et ne peut être substitué par *notamment* et *en particulier*. Dans ce cas on infère *Élaboration**. On propose un axiome 5.4 page 165, sans distinction entre *Élaboration* et *Élaboration**.

Nos analyses ont mis en évidence trois marqueurs toujours spécificateurs de la relation d'*Élaboration* : *notamment*, *en particulier* et *plus précisément* en position initiale détachée. Ces trois adverbes permettent de spécifier la relation, qui est par ailleurs inférable sans leur présence. Leur analyse a également fait surgir des points communs dans leur mode de fonctionnement :

Dans une forme $\langle X$. Notamment/En particulier, $Y \rangle$, soit e_1 les éventualités décrites dans X et e_2 la ou les éventualités décrites dans Y	La ou les éventualités e_2 décrites en Y posent l'existence d'un e_1 et sous-entendent l'existence d'un paradigme d'éventualités plus large Q contenant e_1 , tel que : $e_2 \subset e_1$
Dans une forme $\langle X$. Plus précisément, $Y \rangle$, soit e_1 la ou les éventualités décrites dans X et e_2 la ou les éventualités décrites dans Y :	La ou les éventualités e_2 décrites en Y posent l'existence d'un e_1 et sous-entendent l'existence d'un paradigme d'éventualités plus large Q contenant e_1 , tel que : $e_2 \subseteq e_1$

5.4 Conclusion

Les quatre adverbess paradigmatissants et focalisateurs que nous avons étudiés *notamment*, *en particulier*, *particulièrement* et *précisément* apparaissent fréquemment dans des contextes d'*Élaboration*.

En position intraphrastique, seuls *notamment* et *en particulier* constituent des indices fiables de la relation d'*Élaboration*. Nous avons distingué deux cas de figure présentés dans le tableau ci-dessous :

Paradigme explicite	π_3 Élaboration $\pi_{\{Notamment\}}$
Paradigme implicite	π_3 Autre RD subordonnante $\pi_{\{Notamment/En particulier\}}$ π_3 ————— $\pi_{\{Notamment/En particulier\}}$ Autre RD coordonnante

Nous pouvons considérer qu'ils ont un rôle d'adverbe conjonctif dès lors que la mention du paradigme qu'ils présupposent est explicite. Ainsi l'interprétation du segment introduit par le paradigmatissant s'opère nécessairement en lien avec ce segment du contexte gauche.

En position initiale détachée, nous notons une évolution vers des emplois d'adverbe conjonctif. Nous proposons le tableau ci-dessous pour récapituler les fonctions de nos adverbess dans la mise en œuvre de la relation d'*Élaboration*.

Type d' <i>Élaboration</i>	Marqueurs possibles	Schéma
Isoler un élément comme exemplaire d'un paradigme	<i>notamment</i>	π_a Élaboration $\pi_{\{Notamment\}}$
Isoler un élément comme exemplaire d'un paradigme gradué	<i>plus particulièrement, en particulier</i>	π_a Élaboration $\pi_{\{Plus particulièrement/En particulier\}}$
Opposer une information générale et une information particulière	<i>en particulier, plus particulièrement</i>	$\pi_{[D'une manière générale, En général]}$ Élaboration $\pi_{[En particulier/Plus particulièrement]}$
Mettre en relation deux éléments d'un même paradigme	<i>plus particulièrement</i>	π_a / \ π_b $\pi_{\{Plus particulièrement\}}$
Redescription d'une éventualité	<i>plus précisément</i>	π_a Élaboration $\pi_{\{Plus précisément\}}$
Correction métalinguistique	<i>plus précisément</i>	π_a Élaboration* $\pi_{\{Plus précisément\}}$

Dans ce chapitre, nous avons indirectement illustré le risque d'assigner une fonction discursive unique à un marqueur du discours. Comme nous l'avons vu, nos quatre adverbes apparaissent fréquemment dans des contextes d'*Élaboration* mais pour assigner une fonction discursive à nos adverbes, il faut également tenir compte de la position de celui-ci, de la présence d'autres marqueurs du discours et du co-texte d'apparition. Nous reviendrons sur ces questions dans le chapitre 9.

Chapitre 6

Des marqueurs structurels aux structures énumératives

«Et le deuxièmement, Sir ? s'enquit Yong.

- Hein ? s'éveilla Wadkins.

- Vous avez dit : premièrement. Qu'est ce que vous n'aimez pas, deuxièmement ?

(L'homme chauve-souris, Jo Nesbø, Traduit du norvégien par Élisabeth Tangen et Alexis Fouillet, 2002)

Sommaire

6.1	Marqueurs d'ouverture	170
6.1.1	Double attachement	171
6.1.1.1	Premièrement	171
6.1.1.2	D'abord	171
6.1.1.3	Dans un premier temps	172
6.1.2	Attachement au contexte gauche	174
6.1.2.1	Premièrement	174
6.1.2.2	D'abord	176
6.1.2.3	Dans un premier temps	177
6.1.3	Synthèse	179
6.2	Structures énumératives	181
6.2.1	Les structures énumératives dans les théories du discours	181
6.2.2	Représentation en SDRT des structures énumératives	182
6.2.2.1	Relation entre l'amorce et l'énumération	182
6.2.2.2	Marqueurs d'item comme introducteurs de cadre de dis- cours	190
6.2.2.3	Élaboration des topiques	192
6.2.3	Synthèse	193
6.3	Structures énumératives à deux temps	194
6.3.1	Description des structures énumératives à deux temps	194
6.3.2	Des structures similaires dans les théories du discours	196
6.3.3	Traitement des structures énumératives à deux temps	199
6.3.3.1	Structure de surface	199
6.3.3.2	Structure sémantique profonde	200

6.3.3.3	Illustration	201
6.3.3.4	Généralisation	204
6.3.4	Synthèse	205
6.4	Conclusion	205

Pour poursuivre notre recherche de marqueurs de la relation d'*Élaboration*, après avoir étudié le rôle d'un premier ensemble de marqueurs dans la mise en œuvre de cette relation, nous allons, dans ce chapitre, proposer une étude d'une configuration d'indices au sein d'une structure textuelle, appelée *structure énumérative*. De façon intuitive, il est possible de penser que les structures énumératives sont des structures isomorphes à des structures d'*Élaboration*. Mais nous allons montrer dans ce chapitre que cet isomorphisme n'est pas si évident.

Nous commençons par présenter des études sur quelques marqueurs appelés *d'ouverture* (Bras, 2008), *d'abord*, *dans un premier temps* et *premièrement*, qui apparaissent généralement dans des structures énumératives. De ces travaux, nous allons montrer une gradation des emplois temporels aux emplois textuels et l'implication de cette gradation dans la mise en œuvre de la relation d'*Élaboration*. Avec pour objectif de poursuivre notre inventaire des marqueurs de la relation d'*Élaboration*, nous allons également montrer qu'il n'est pas possible d'associer à chacun de ces marqueurs une fonction rhétorique unique. Nous montrerons que bien qu'ils apparaissent incontestablement dans des contextes d'élaboration, les marqueurs présentent des fonctionnements différents. Ces marqueurs apparaissent fréquemment dans des structures énumératives qui font l'objet de notre deuxième section. De ce fait, nous allons nous intéresser davantage à leur rôle textuel et organisationnel. Nous présenterons les structures énumératives avec pour objectif de déterminer le rôle de chacun des éléments constitutifs de celles-ci dans la mise en œuvre de la relation d'*Élaboration* et de la relation d'*Énumération* que nous proposons d'introduire dans le cadre de la SDRT pour rendre compte des informations affleurant à la surface des textes. Étudier les structures énumératives, qui ont par ailleurs fait l'objet de nombreuses études, notamment dans le cadre de la linguistique textuelle (Adam et Revaz, 1989), du point de vue des relations de discours à l'œuvre permet de porter un regard neuf et d'approfondir nos connaissances sur ces structures. Enfin, nous terminons ce chapitre avec l'étude des structures énumératives à deux temps, qui sont un type particulier de structures énumératives. Nous mettons ainsi à l'épreuve le traitement proposé pour les structures énumératives sur ces structures énumératives à deux temps et montrons que ces structures constituent un défi pour les théories du discours car elles mettent en œuvre des mécanismes qui semblent, à première vue, capables de traverser la frontière droite.

6.1 Marqueurs d'ouverture

Les adverbes *d'abord*, *dans un premier temps*, *premièrement*, *en premier lieu* sont tous des adverbes classifiés comme *marqueur d'intégration linéaire* (Turco et Coltier, 1988), comme un sous-type de marqueurs du discours que nous appelons marqueurs d'ouverture. Ces marqueurs d'intégration linéaire constituent, selon ces auteurs, une classe linguistique de marqueurs qui «accompagnent l'énumération sans fournir de précision autre que le fait que le segment discursif qu'ils introduisent est à introduire de façon linéaire dans la série»

(Turco et Coltier, 1988, p. 57). Dans cette classe, nos adverbes sont plus précisément considérés comme des marqueurs d'ouverture d'une série. Les études présentées ci-dessous, s'appuyant sur (Bras, 2007; Bras *et al.*, 2008; Bras et Le Draoulec, 2009) pour *d'abord*, sur (Bras et Schnedecker, 2009) pour *dans un premier temps* et sur (Schnedecker, 2001a) pour *premièrement*, proposent une analyse plus approfondie de ces adverbes et mettent en évidence des différences dans leur fonctionnement discursif.

6.1.1 Double attachement

Un principe de double attachement a été décrit pour chacun de ces trois adverbes.

6.1.1.1 *Premièrement*

En s'appuyant sur les travaux de Guimier (1996), Molinier et Lévrier (2000) et de Nøjgaard (1995), Schnedecker (2001a) note que *premièrement* requiert un contexte gauche et un contexte droit. Elle décrit sa contribution comme le déclenchement d'un double mécanisme d'anticipation et de rétroaction.

- (1) Je crois que j'avais fait en même temps trois découvertes : **premièrement** les adultes avaient un «travail», **deuxièmement** ce travail des adultes était une chose capitale, auprès de laquelle rien de ce que j'avais appris auparavant ne comptait, et **troisièmement**, j'ignorais quelle était cette chose capitale en ce qui concernait notre père. (Guimier, 1996, p. 126)

Le principe d'attachement au contexte droit est illustré avec l'épigraphe du début du chapitre : Wadkins a commencé une série qu'il a introduit avec *premièrement* et Yong est dans l'expectative d'un *deuxièmement*.

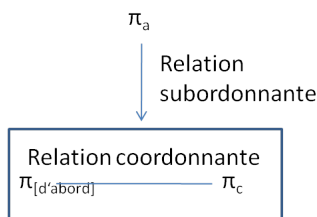
6.1.1.2 *D'abord*

De même pour *d'abord*, celui-ci est décrit d'une part comme un adverbial conjonctif : le segment de discours qu'il introduit est d'une part lié à un segment de discours situé dans le contexte gauche (Guimier, 1996; Molinier et Lévrier, 2000) et d'autre part comme un marqueur d'ouverture d'une série située dans le contexte droit (Turco et Coltier, 1988; Adam et Revaz, 1989; Nøjgaard, 1995; Guimier, 1996). Cette propriété de double attachement discursif a motivé l'analyse de Bras (2007) en termes de relations coordonnante et subordonnante dans le cadre de la SDRT que nous allons présenter ci-dessous.

D'abord, en tant qu'adverbial conjonctif, établit un lien avec son contexte gauche. Bras (2007) montre que le constituant contenant l'adverbe est attaché, par une relation de discours subordonnante, à un constituant considéré comme topique de discours.

D'abord, en tant que marqueur d'ouverture d'une série, introduit le premier élément d'une série. De plus il annonce qu'il sera suivi par, au moins, un autre élément. Cet élément sera attaché par une relation coordonnante comme *Narration* ou *Continuation*. Ce double attachement est schématisé Figure 6.1 et illustré avec l'exemple (2).

- (2) [Une campagne de dénigrement, trop systématique pour n'être pas organisée, se déchaîna contre la personne même de Sélim. (...)]_a [**D'abord** on répandit le bruit absurde que Sélim était alcoolique, alors qu'il était la sobriété même.]_b [Puis on

FIG. 6.1 – Principe du double attachement de *d'abord*

prétendit partout que Sélim était stérile, c'est-à-dire maudit.]]_c (Michel de Grèce, La nuit du sérail, 1982, p.373)

En (2), le segment (b) élabore (a), qui a ici le rôle de topique de discours explicite. Le segment (c) est relié au segment (b) par une relation de *Narration*, inférée grâce à la présence du connecteur *puis*. Par le principe de distributivité, (c) poursuivant la fonction rhétorique de (b), élabore (a). La relation d'*Élaboration* est ici inférée car le prédicat *Subtype_D* est vérifié grâce à un ensemble d'informations alliant connaissances du monde et sémantique lexicale, exprimant que le type de l'événement décrit en (b) «répandre le bruit que Sélim était alcoolique» est un sous-type de celui décrit en (a) «déchaîner une campagne de dénigrement contre Sélim». On peut donc inférer *Élaboration* sans tenir compte de la présence de *d'abord*, qui peut par ailleurs être supprimé sans affecter la cohérence du discours.

Comme nous l'avons vu dans le chapitre 4, les marqueurs du discours ont très souvent vocation à faciliter le travail interprétatif du lecteur. *D'abord* facilite l'interprétation de (b) comme une première étape de la «campagne de dénigrement» et prépare l'interprétation d'un deuxième élément. Pour revenir à l'épigraphe de ce chapitre, nous sommes dans la même situation où le protagoniste Yong est dans l'attente d'un *deuxièmement*. Le discours en (2) perd de sa cohérence si l'on supprime le segment (c).

- (3) ?? [Une campagne de dénigrement, trop systématique pour n'être pas organisée, se déchaîna contre la personne même de Sélim. (...)]_a [**D'abord** on répandit le bruit absurde que Sélim était alcoolique, alors qu'il était la sobriété même.]_b

Le lecteur se prépare à interpréter une autre étape de la campagne de dénigrement. Dans cette configuration, *d'abord* introduit une première étape d'élaboration mais nous allons voir que l'adverbe est compatible avec d'autres relations de discours subordonnantes.

6.1.1.3 Dans un premier temps

Dans un premier temps est classifié comme un adverbial ordinal avec d'autres adverbes tels que *premièrement*, *en premier lieu*... Ces adverbes sont formés sur la base d'un adjectif ordinal. Ils sont ainsi considérés comme signalant un rang dans une série (Guimier, 1996). *Dans un premier temps* fait aussi partie des marqueurs d'intégration linéaire en tant que marqueur d'ouverture d'une série. Nøjgaard (1995) classe *dans un premier temps* dans les relationnels sériels, mettant en avant l'interdépendance de ce marqueur. Par exemple,

après l'occurrence de *dans un premier temps* en (4) (Bras et Schnedecker, 2009)¹, une suite décrivant un second temps est attendue.

- (4) ?? Le député de Toul étant mort, une élection partielle est organisée. **Dans un premier temps**, la seule candidature républicaine est celle d'un modéré, qui se prononce contre l'amnistie. (Frantext : Hugo, Depuis l'exil, 1885) (Bras et Schnedecker, 2009)

Bras et Schnedecker (2009) signalent que les études précédemment citées (Guimier, 1996; Turco et Coltier, 1988; Nøjgaard, 1995) sur *dans un premier temps* ne rendent pas compte efficacement du rôle et de la fonction de *dans un premier temps*. D'abord, celles-ci ne rendent pas compte de la diversité lexicale des adverbiaux ordinaux. Ensuite, elles ne rendent pas compte de la diversité fonctionnelle des éléments de cette classe qui ne peuvent pas tous permuter dans le même contexte.

- (5) Le député de Toul étant mort, une élection partielle est organisée.
***D'abord/*Premièrement/*En premier lieu**, la seule candidature républicaine est celle d'un modéré, qui se prononce contre l'amnistie. (Bras et Schnedecker, 2009)

Enfin, elles ne distinguent pas les diverses fonctions et valeurs que chaque marqueur peut véhiculer, par exemple ordonner des événements comme en (6) ou ordonner des segments discursifs comme en (7).

- (6) La physiologie des glandes mammaires évolue en deux temps. **Dans un premier temps** les glandes se développent pour être prêtes à assurer la production de lait. **Dans un deuxième temps**, facultatif celui-là, se produit la sécrétion lactée, au décours des accouchements (Frantext : Encyclopédie Quillet, 1965) (Bras et Schnedecker, 2009)
- (7) Ma thèse s'articule autour de trois grands axes. **Dans un premier temps**, je m'intéresse aux patrons de dynamique en métapopulation en couplant l'utilisation de modèles démographiques et l'analyse des flux de gènes. **En second lieu**, j'analyserai les facteurs influant sur les processus de colonisation et d'extinction, notamment à la lumière de la densité des populations. **Enfin, en dernier lieu**, je chercherai à déterminer quel rôle jouent les structures florales sur l'attraction des pollinisateurs. (Web) (Bras et Schnedecker, 2009)

Comme *premièrement* et *d'abord*, les emplois discursifs de *dans un premier temps* font intervenir des relations avec son contexte gauche et son contexte droit. Ces emplois sont analysés par Bras et Schnedecker (2009) dans le cadre de la SDRT. L'attachement au contexte gauche et droit peut être illustré par un cas typique de l'emploi de *dans un premier temps* donné en (6) et dont la structure discursive est donnée en 6.2.

(b) est le segment introduit par *dans un premier temps*. (a) est le segment cible auquel est attaché le segment (a). Et (c) est le segment attendu dans le contexte droit.

Le segment (a) décrit une éventualité temporellement hétérogène «évoluer en deux temps». L'expression «en deux temps» déclenche à la fois une interprétation temporelle et

¹Ce travail a d'abord fait l'objet d'une communication lors de la conférence LPTS. Les exemples et leurs analyses sont issus de la version papier (Bras et Schnedecker, à paraître).

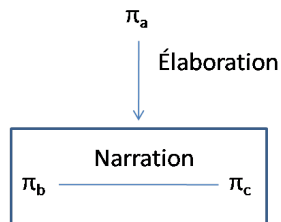


FIG. 6.2 – Représentation discursive de (6)

de manière. Les segments (b) et (c) élaborent le contenu sémantique décrit dans le segment (a) en détaillant les deux temps de l'évolution. La relation ici à l'oeuvre est une relation de *partie de* (*Subevent_D*).

6.1.2 Attachement au contexte gauche

En (1) pour *premièrement*, (2) pour *d'abord* et (6) pour *dans un premier temps*, la relation de discours inférée pour le rattachement du segment contenant nos marqueurs d'ouverture au contexte gauche est *Élaboration*. En ce sens, ils sont pertinents pour notre étude mais nous allons voir qu'ils ne sont pas des marqueurs de la relation d'*Élaboration*.

6.1.2.1 *Premièrement*

Des marqueurs structurels que nous avons présentés jusqu'à maintenant dans ce chapitre, *premièrement* est celui qui semble le mieux s'adapter à la définition des marqueurs d'intégration linéaire.

Cette fonction des marqueurs d'intégration linéaire s'applique fortement à l'adverbe *premièrement* du fait que celui-ci, ayant pour base morphosémantique la numération, élude toute référence explicite au temps (contrairement à *dans un premier temps*) et à l'espace (contrairement à *en premier lieu*). Schnedecker (2001a) signale que la longueur et la position initiale stratégique de ces adverbes ordinaux (*premièrement*, *deuxièmement*, *troisièmement*) rend ces adverbes saillants dans les textes et leur confère un rôle de formation, au niveau visuel, de blocs textuels. Nous verrons également, au niveau sémantique et discursif, leur rôle d'introducteur de cadre suivant les hypothèses de Charolles (1997) dans la section 6.2.2.2.

Les adverbes ordinaux ont une fonction paradigmatitante, tout comme les adjectifs dont ils découlent. Ils présupposent un ensemble fermé dont ils délimitent un sous-ensemble.

En discours, l'ensemble fermé présupposé par l'emploi des adverbiaux ordinaux est souvent mentionné au sein d'une amorce, comme en (1), avec «Je crois que j'avais fait en même temps trois découvertes». Le SN «trois découvertes» fournit des informations sur cet ensemble. Nous reviendrons sur ces structures et leur traitement dans la section sur les structures énumératives (section 6.2).

En l'absence de cette amorce, Schnedecker (2001a) précise les différentes fonctions possibles de l'énoncé qui introduit la série dans un contexte monologal ou dialogal. Nous

allons nous appuyer sur ses données et préciser la relation de discours à l'œuvre entre l'énoncé introducteur et la série introduite par *premièrement* pour harmoniser les descriptions de Schnedecker (2001a) avec les formalisations proposées pour *d'abord* et *dans un premier temps*.

Nous nous restreindrons au contexte monologal. Schnedecker (2001a) signale qu'en (8), la phrase introductrice fait office de thèse. La série introduite par *premièrement* est alors argumentale.

- (8) Tous les Guadeloupéens sont parents. **Premièrement**, ils sont pour la plupart sortis du même ventre-négrier, expulsés au même moment, sur les mêmes marchés aux esclaves. **Deuxièmement**, dans les plantations, des liens se sont noués entre ceux-là et les autres, promiscueux, proches comme des incestes. (M. Condé, roman) (Schnedecker, 2001a, p. 275)

Dans le cadre de la SDRT, nous inférons la relation d'*Explication* en nous basant sur les liens de causalité entre les types des éventualités. «Les guadeloupéens viennent des mêmes marchés aux esclaves» (deuxième phrase) et «les guadeloupéens ont noués des liens entre eux» (troisième phrase) sont des causes du type de l'éventualité «tous les guadeloupéens sont parents» (première phrase).

Premièrement initie un décrochement. Il démarre une nouvelle fonction rhétorique. Schnedecker (2001a) décrit toutes les séquences introduites par les adverbiaux ordinaux comme se situant à un même niveau et comme poursuivant la même fonction rhétorique vis-à-vis du segment introducteur. Elle donne un exemple de deux séquences hors contexte (Schnedecker, 2001a, p. 277) :

- (9) Il vit à Walthamstow, autant dire à Pétaouchnock.
(10) Il se prénomme Neville.

La valeur argumentative de ces énoncés n'est pas claire hors contexte. En revanche, le contenu d'une phrase introductrice peut orienter une interprétation :

- (11) Il y a tout de même un ou deux petits détails qui me chiffonnent. **Premièrement**, il vit à Walthamstow, autant dire à Pétaouchnock. **Deuxièmement**, il se prénomme Neville.

Les deux séquences introduites sont ainsi comprises comme deux arguments en défaveur du personnage en question. Dans ce cas, nous inférons la relation d'*Élaboration*. Les deux types des éventualités «il vit à Walthamstow» et «il se prénomme Neville» sont des sous-types du type «il existe des détails qui me chiffonnent».

Ce fonctionnement décrit à travers le marqueur spécifique *premièrement*, constitue de notre point de vue, le fondement même de la structure énumérative, quelque soit les marqueurs d'intégration linéaire employés pour introduire les séquences. Nous reviendrons sur ces questions dans la section 6.2.

Une étude plus approfondie des contextes d'apparition en discours de l'adverbe permettrait de mettre au jour les relations pouvant apparaître entre la phrase introductrice et la série.

6.1.2.2 *D'abord*

L'attachement au contexte gauche du segment contenant *d'abord* se fait toujours par une des relations subordonnantes suivantes : *Élaboration* comme en (2), *Explication* comme en (12) et (13), *Résultat* comme en (14) et *Précondition* (ou *Flashback*) comme en (15).

- (12) [La discipline essentielle est alors le latin,]_a [**d'abord** parce que le latin est la langue de l'église,]_b [parce qu'aussi c'est à Rome plus qu'à Athènes qu'on cherche ses maîtres à penser.]_c (Encyclopédie pratique de l'éducation en France, 1960, p. 125)
- (13) [Tu devrais épouser Laura, Jean-Pierre.]_a [**D'abord**, elle est adorable.]_b [Ensuite, c'est une des plus riches héritières du Brésil.]_c (Romain Gary, Au-delà de cette limite votre ticket n'est plus valable, 1975, p. 253)

Cet usage de *d'abord* est très fréquent, que la relation d'*Explication* soit explicitement marquée ou non. Le segment (c) des deux exemples, en tant que deuxième élément de l'explication, est attaché au segment (b) par la relation de *Continuation*. La présence de *d'abord*, là aussi, prépare l'interprétation de la seconde partie de l'explication. On ne note pas une corrélation entre *d'abord* et la relation d'*Explication*, mais *d'abord* permet d'avoir une explication double.

- (14) Une nuit, [il lui a demandé de faire la même chose pour moi <il lui a demandé de raconter mon enfance> (...)]_a [**D'abord**, Mariella a fait une scène de jalousie,]_b [puis elle a répondu qu'il fallait qu'elle me demande l'autorisation de dévoiler mon passé.]_c (Evane Hanska, Les Amants foudroyés, 1984, p. 62)

Le segment (b) décrit une réponse ou une réaction à la demande en (a). *D'abord* introduit un premier élément de réponse et prépare l'interprétation d'un deuxième élément de réponse. Le segment (c) introduit par *puis* est attaché au segment (b) par *Narration* et au segment (a) par *Résultat*².

- (15) [Elle avait froid]_a [**D'abord**, elle avait eu trop chaud quand le liquide de Maman-Madeleine avait coulé sur elle.]_b [Mais maintenant la mouillure refroidissait]_c (F. Chandernagor, L'enfant des lumières, 1995) (Bras et Le Draoulec, 2009, p. 240)

La relation de *Précondition* est inférée lorsqu'un segment introduit une éventualité qui correspond à un retour en arrière dans la narration. Comme le soulignent les auteurs, cette relation est particulièrement pertinente pour décrire certains emplois de *d'abord*.

Bras et Le Draoulec (2009) signalent que, dans aucun exemple attesté de leur corpus, le constituant introduit par *d'abord* n'est attaché au contexte gauche par la relation d'*Arrière-plan* et qu'il est difficile de construire un exemple d'une telle configuration.

L'analyse de Bras (2007) montre le rôle facilitateur de *d'abord* dans la mise en œuvre d'une relation subordonnante. Dans la plupart des cas, la relation de discours est inférable en dehors de l'action de *d'abord*. En s'appuyant sur des exemples de Péroz (2001), elle

²Les segments (b) et (c) constituent intuitivement un segment complexe, noté π_A , constitué d'une réaction en deux temps, lié au segment (a) par *Résultat*. Le résultat ou la réaction en deux temps entraîne ainsi que la relation *Résultat* soit subordonnante, en suivant les propositions d'Asher et Vieu (2005).

met en évidence qu'il existe des cas où *d'abord* est nécessaire pour inférer une relation subordonnante :

(16) [On a mangé sur la terrasse.]_a [Mon père est arrivé.]_b

(17) [On a mangé sur la terrasse.]_a [**D'abord** mon père est arrivé.]_b

Péroz indique qu'on interprète les deux segments (a) et (b) en (16) comme une succession de procès considérés sur le même plan discursif tandis que (17) institue le segment (b) dans une relation de dépendance avec (a). L'analyse de Bras (2007) va dans le même sens : une relation coordonnante, *Narration*, peut être inférée entre les segments (a) et (b) en (16) ; en (17), le segment (b) ne fait pas avancer le récit et *d'abord* nécessite de revenir en arrière sur le segment (a) pour en élaborer sa description, par exemple en commençant à décrire les participants du repas.

Bras traduit ce rôle de *d'abord* par l'introduction de la relation subordonnante minimale, notée \Downarrow . La relation imposée par *d'abord* ne peut pas être *Élaboration* car celle-ci doit rester compatible avec d'autres relations de discours subordonnantes inférables à partir d'autres indices disponibles dans le texte, comme *Explication* en (12) et (13), *Résultat* en (14) ou *Précondition* en (15). Bras et Le Draoulec (2009) proposent l'axiome suivant :

Axiome 6.1 *Inférer_subord_coord_d'abord_bis* $?(α, β, λ) \wedge [d'abord]β > (\exists R_1 \in \text{Subord} - \{\text{Arrièreplan}\})R1 (α, β, λ) \wedge \exists \delta \exists R_2 \in \text{Coord } R2(β, \delta, λ)$

Cet axiome permet de bloquer l'inférence d'*Arrière-plan* et d'inférer une relation subordonnante sous-spécifiée lorsqu'aucune autre relation de discours subordonnante n'est vérifiée.

Le rôle premier de *d'abord* semble donc d'instaurer cette relation subordonnante minimale en préparant l'interprétation d'(au moins) un deuxième segment (δ dans l'axiome) qui sera attaché, à celui contenant *d'abord*, par une relation coordonnante.

D'abord en tant que marqueur d'une relation subordonnante est un indice intéressant pour le repérage de la relation d'*Élaboration*.

Le même type d'étude a été proposé par Bras et Schnedecker (2009) pour le marqueur *dans un premier temps*, étude que nous résumons dans la section suivante.

6.1.2.3 Dans un premier temps

Dans un premier temps a besoin pour être interprété d'accéder à une éventualité globale hétérogène dont il découpe une partie. Le fonctionnement de *dans un premier temps* à l'œuvre avec le contexte gauche varie selon que cette éventualité globale est explicite ou non.

Dans des cas tels que (6) et (7)³, Bras et Schnedecker (2009) introduisent le terme d'*Élaboration de topique explicite*. L'amorce (phrase introductrice) de ces structures annonce la nature hétérogène de l'éventualité dénotée, comme «évolution en deux temps».

³Ces exemples sont par ailleurs des cas typiques de structures énumératives que nous étudierons plus en détail dans la section 6.2.

En (18), *dans un premier temps* introduit une partie d' une éventualité globale qui n'est pas donnée dans le contexte gauche.

- (18) [Mais à partir de ce foyer colique il arrive que des colonies amibiennes parviennent par voie sanguine au niveau du foie.]_a [**Dans un premier temps** il se produit autour des intrus une réaction inflammatoire hépatique]_b [**puis** sous l'afflux de globules blancs se forme une poche de pus]_c (Frantext : Encyclopédie Quillet, 1965) (Bras et Schnedecker, 2009)

Cette éventualité globale peut être paraphrasée comme «ce qui s'est passé à l'arrivée des colonies amibiennes au niveau du foie». Les segments (b) et (c) décrivent en deux temps le résultat de l'éventualité dénotée en (a). Bras et Schnedecker (2009) décrivent ces configurations comme des *Élaborations de topique implicite*. Cette éventualité globale est introduit dans la structure au moyen d'un constituant implicite, noté π^* (Cf. Figure 6.3).

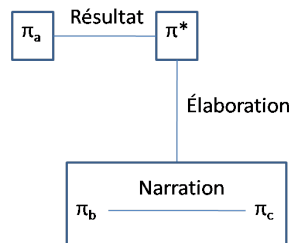


FIG. 6.3 – Représentation discursive de (18)

Bras et Schnedecker (2009), contrairement au traitement proposé pour *d'abord*, proposent que le marqueur *dans un premier temps* nécessite la création de ce topique implicite qui doit inclure l'éventualité globale développée en plusieurs étapes. Ce traitement s'appuie sur la différence de fonctionnement des deux adverbes *d'abord* et *dans un premier temps*. L'emploi de *dans un premier temps* nécessite l'introduction d'une entité temporelle hétérogène localisée par l'adverbe, contrairement à *d'abord* qui n'a pas cette fonction de localisation.

D'autres configurations d'*Élaboration de topique implicite* mettent en jeu les relations subordonnantes d'*Explication* en (19) et de *Précondition* (ou *Flashback*) en (20).

- (19) [Pour le match inaugural du 28 janvier, la Fédération française de football a monté un chapiteau à 500 mètres du Stade de France.]_{a1} [Pour le France-Angleterre du 7 février, la Fédération française de rugby projette de faire de même.]_{a2} [**Dans un premier temps**, les deux fédérations avaient pourtant eu dans l'idée d'installer leurs traditionnelles agapes dans la salle polyvalente du Stade de France.]_b [**mais** elles ont reculé en apprenant les tarifs de location.]_c (Le Monde) (Bras et Schnedecker, 2009)
- (20) [Michel Barnier ne fait plus partie du conseil d'administration du parc national de la Vanoise, (...).] [La ministre de l'environnement, (...) lui a en effet préféré Thierry Repentin, adjoint (PS) au maire de Chambéry et conseiller général.]_a [**Dans un premier temps**, ce dernier avait souhaité être désigné par le conseil général.]_b

[**mais** M. Barnier s'y était opposé, préférant que les trois sièges dont dispose l'assemblée départementale soient détenus par des élus de sa majorité]_c (Le Monde) (Bras et Schnedecker, 2009)

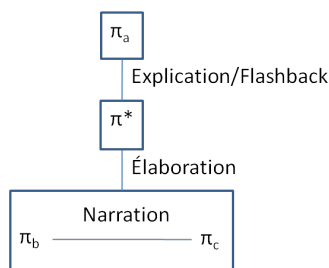


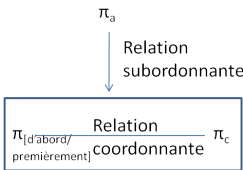
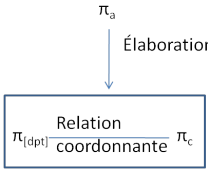
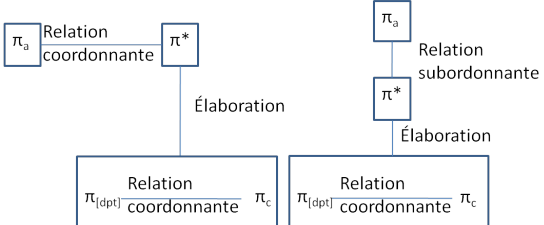
FIG. 6.4 – Représentation discursive de (19) et (20)

Dans un premier temps signale une *Élaboration*. Celle-ci est établie soit avec un topique explicite, soit avec un topique implicite. Nous le considérons comme un indice intéressant pour le repérage de la relation d'*Élaboration*. L'étude de la fonction du marqueur dans la mise en œuvre de la relation reste encore à être établie.

6.1.3 Synthèse

Premièrement, de par sa fonction de numération, sans référence explicite au temps ou à l'espace, nous paraît le mieux correspondre à la définition des marqueurs d'intégration linéaire. *Dans un premier temps* est celui qui correspond le moins à la définition. Dans le corpus de Bras et Schnedecker (2009), aucune occurrence de *dans un premier temps* ne joue un rôle exclusif d'organisation de la matière textuelle. On peut trouver ce type d'occurrence dans des genres très spécifiques comme les thèses, les mémoires, les articles scientifiques dans lesquels ce marqueur semble dérouler le temps de l'exposition d'un plan. *D'abord* se situe entre les deux pouvant jouer un rôle au niveau de l'ordre des éventualités mais également de l'ordre des segments discursifs. Enfin *en premier lieu*, contrairement à *dans un premier temps* qui a conservé son interprétation temporelle, semble avoir perdu son interprétation spatiale au profit d'une interprétation de marqueur d'intégration linéaire.

Du point de vue des relations de discours, le rôle de *d'abord* n'est pas de «démarrer» une élaboration mais de «démarrer» une nouvelle fonction rhétorique. Le rôle de *dans un premier temps* semble, au contraire, d'imposer une relation d'*Élaboration* mais cette relation est inférée dans certains cas avec un topique explicite et dans les autres avec un topique implicite. Enfin, *premièrement* dont les fonctions discursives n'ont pas fait l'objet d'une étude détaillée dans le cadre de la SDRT, semble avoir un rôle plus similaire à *d'abord* qu'à *dans un premier temps*.

Relations de discours	Marqueurs possibles	Schéma
Relations subordonnantes	<i>d'abord, premièrement</i>	
<i>Élaboration d'un topique explicite</i>	<i>dans un premier temps</i>	
<i>Élaboration d'un topique implicite</i>	<i>dans un premier temps</i>	

TAB. 6.1 – Synthèse des emplois discursifs des marqueurs d’ouverture

Nous allons poursuivre l'analyse de ces marqueurs en considérant leurs emplois dans les structures énumératives.

6.2 Structures énumératives

Les marqueurs d'ouverture tels que *d'abord*, *dans un premier temps*, *en premier lieu* et *premièrement* apparaissent très fréquemment dans des structures énumératives. Intuitivement, il semblerait que ces structures soient isomorphes à des structures d'élaboration.

Prenons l'exemple (21) pour illustrer les parties constitutives d'une structure énumérative :

- (21) [Domenech refuse obstinément cette concession pour **deux raisons**.]_a [**D'abord**, il ne la comprend pas.]_b [**Ensuite**, en l'acceptant, il aurait le sentiment de ruiner une autorité déjà amplement chancelante.]_c (Réponse à tout)

(21) peut être décrit comme une structure énumérative faisant apparaître une amorce «Domenech refuse obstinément cette concession pour deux raisons», segment qui annonce une énumération. Cette amorce peut contenir un SN *prospectif* (ou prospect), qui est un SN lexical à valeur cataphorique, ici «deux raisons». Il se compose, dans ce cas, d'un numéral «deux», qui précise le nombre d'items et d'un *énumérathème*⁴ ou *classificateur énumératif* (Schnedecker, 2001a) «raisons» qui spécifie le critère de co-énumérabilité. Les prospects peuvent également être introduits par des indéfinis *plusieurs*, *quelques...* ou des collectifs *une foule de*, *une grande variété de...* L'amorce permet d'introduire une énumération sur des événements, des états ou des entités.

Enfin cette énumération est constituée de deux items (b) et (c).

Les structures énumératives ont souvent été évoquées implicitement dans les théories du discours et bien qu'elles aient été beaucoup étudiées, ont peu fait l'objet d'analyse en relations de discours (sous-section 6.2.1). Nous allons proposer un traitement, dans le cadre de la SDRT, des structures énumératives : d'abord en nous interrogeant sur la ou les relations s'établissant entre l'amorce et l'énumération (section 6.2.2.1), puis en proposant une analyse discursive du rôle des marqueurs d'item (section 6.2.2.2). Nous terminons en proposant une discussion sur l'*Élaboration* des topiques dans les structures énumératives (section 6.2.2.3).

6.2.1 Les structures énumératives dans les théories du discours

Les structures énumératives, ou plus précisément les énumérations apparaissent à travers la relation de *List* de la RST ou les séquences présentationnelles et les séquences temporelles de Knott (1996). Ces derniers identifient, dans leur liste de marqueurs de relations, des marqueurs de début de séquence, de position dans une séquence, d'étape suivante, *etc.* Ces marqueurs et/ou relations sont peu illustrés et peu discutés, du fait que Mann et Thompson (1987) et Knott (1996) cherchent plutôt à recenser un ensemble de relations de discours plutôt qu'à détailler les particularités de chacune.

⁴Les termes *SN prospectif* et *énumérathème* ont été définis dans le cadre du projet ANNODIS par Ho-Dac et al. (2009).

Le modèle de Polanyi (1988) ne traite pas spécifiquement, à notre connaissance, de structures énumératives, mais les structures N-aires récemment introduites (Polanyi et al, 2004) nous semblent intéressantes pour analyser ces dernières. Les structures N-aires sont une sorte de généralisation des relations binaires incluant principalement des paires antécédent/conséquent ou des paires dialogiques comme question/réponse. L'extension à des structures N-aires est justifiée d'une part par l'existence de structures pragmatiques de plus haut niveau et d'autre part par la découverte de marqueurs signalant de telles structures, comme par exemple les adverbes en position initiale. Cependant, bien qu'ils identifient ces structures, Polanyi et ses collègues ne proposent ni une analyse fine des marques de surfaces les signalant, ni une analyse linguistique de leurs conséquences pour l'interprétation du discours. Ils notent cependant une différence entre les constructions coordonnantes et les constructions N-aires. Ces deux structures supposent l'existence d'un constituant commun aux éléments coordonnés ou aux éléments groupés sous la relation N-aire, et auquel ils sont attachés. Dans le cas des relations coordonnantes, ce constituant contient les informations communes à tous les éléments coordonnants. À l'inverse, pour les constructions N-aires, ce constituant contient les informations de chaque élément de la construction ainsi que la relation elle-même.

Le travail de Luc (2000) traite spécifiquement des structures énumératives et de la façon dont on peut combiner le Modèle d'Architecture Textuelle (MAT) qui fournit des informations pour la délimitation des segments complexes en s'appuyant sur la mise en forme matérielle (niveau textuel) et la RST qui fournit les relations entre ces objets textuels, mis au jour par le MAT, et les relations à l'intérieur de ces objets, lorsque ceux-ci sont des segments complexes (niveau sémantique). Afin de prendre en compte les deux modèles, Luc a proposé d'utiliser un formalisme de représentation semblable pour la RST et pour le MAT et dresse entre les deux modèles des liens multiples afin de pouvoir les superposer et les comparer. Ce travail est un des rares à prendre en compte le niveau textuel (via le MAT) et celui des relations de discours en étudiant les correspondances et la complémentarité entre les deux niveaux. Dans ce chapitre, nous avons cherché à prendre en compte de façon conjuguée ces deux dimensions, en examinant la façon dont elles interagissent avec le contenu sémantique des segments, en employant le modèle de la SDRT et en intégrant en son sein des informations d'ordre textuel et architectural.

6.2.2 Représentation en SDRT des structures énumératives

6.2.2.1 Relation entre l'amorce et l'énumération

Bras (2007) distingue les structures énumératives présentant des énumérations temporellement ordonnées comme en (22) des structures non temporellement ordonnées comme en (21).

- (22) [Tout en roulant vers l'île Saint-Louis, il prit deux décisions.]_a [**D'abord**, il allait persuader sa mère de quitter Paris au moins jusqu'au procès et de descendre dans le Midi, peut-être même en Italie où elle ne risquerait rien.]_b [Et puis, de son côté, il abandonnerait la rue de Longchamp et louerait un studio meublé.]_c (Michel Droit, *Le Retour*, 1964)

Dans les structures comme (22), l'interprétation temporelle de l'énumération est préparée

par le contenu du prospect qui contient une expression référentielle visant des référents temporels, syntagmes nominaux à base de noms d'événement ou de noms de temps. Dans ce cas, les structures énumératives sont isomorphes à des structures d'*Élaboration*, Cf. Figure 6.5.

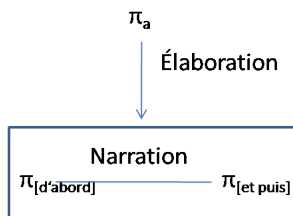


FIG. 6.5 – Structure de (22)

L'énumérathème, spécifiant le critère de co-énumérabilité des items de l'énumération, joue un rôle crucial dans l'inférence des relations de discours : si l'énumérathème est un nom d'événement ou un nom de temps, cela encourage l'interprétation de la structure énumérative comme isomorphe à une structure d'*Élaboration*.

Sinon, pour les cas illustrés en (21), l'énumérathème semble donner des indications sur la relation à l'œuvre entre l'amorce et l'énumération.

Nous allons reprendre l'exemple (21) et discuter les différentes représentations possibles de cet exemple avec les outils proposés dans le cadre de la SDRT.

Solution 1 : Attacher directement le segment (b) au segment (a)

Cette solution revient à traiter l'énumération en (21) comme s'attachant à l'amorce par *Explication*; Cf. Figure 6.6. Le SN prospectif est ainsi considéré comme un connecteur marqueur d'une explication double. Ce connecteur serait équivalent à l'emploi du connecteur *parce que* comme illustré en (23).

- (23) [Domenech refuse obstinément cette concession.]_a. [D'abord parce qu'il ne la comprend pas.]_b [Ensuite parce qu'en l'acceptant, il aurait le sentiment de ruiner une autorité déjà amplement chancelante.]_c

Mais cela n'est évidemment pas le cas. Le rôle d'un connecteur est d'introduire un segment en donnant des instructions sur la façon dont celui-ci est attaché au contexte gauche. Un SN prospectif, tel que «pour deux raisons» ne va pas obligatoirement induire l'introduction d'une énumération, comme en atteste l'exemple suivant :

- (24) [Domenech refuse obstinément cette concession pour deux raisons.]_a [Mais il a préféré ne pas nous les communiquer.]_b

Cette solution ne rend pas fidèlement compte du rôle de l'amorce dans la représentation discursive. Nous voyons, de façon intuitive, que les items de l'énumération sont interprétés sous le critère de co-énumérabilité désigné par l'énumérathème «raisons».

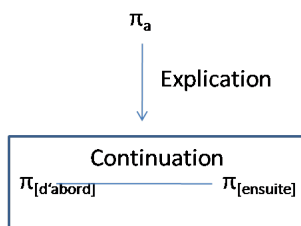


FIG. 6.6 – Structure de (13)

L'amorce joue un rôle primordial, surtout en l'absence de marqueurs d'items.

- (25) Progressivement, ces diversifications ont pris de l'importance dans le chiffre d'affaires, dont elles représentent aujourd'hui 35 %. Et leur part pourrait atteindre les 50 % d'ici à 1994 ou 1995, **pour deux raisons** : la limitation de la production laitière due aux quotas européens et la pénétration d'un nouveau marché, celui du jus de fruits en bocal de verre. (Le Monde)

En (25), elle ne peut être supprimée sans affecter la cohérence du discours :

- (26) *Progressivement, ces diversifications ont pris de l'importance dans le chiffre d'affaires, dont elles représentent aujourd'hui 35 %. Et leur part pourrait atteindre les 50 % d'ici à 1994 ou 1995 : la limitation de la production laitière due aux quotas européens et la pénétration d'un nouveau marché, celui du jus de fruits en bocal de verre.

L'énumérathème «raisons» prépare, en quelque sorte, la relation d'*Explication* : les événements décrits dans les items «la limitation de la production laitière» et «la pénétration d'un nouveau marché» sont des explications de l'augmentation probable du chiffre d'affaires de la compagnie.

Dans cette même configuration d'attachement, nous soulevons la question d'employer la nouvelle relation de discours *Élaboration d'entité*. En (21), les segments (b) et (c) seraient ainsi attachés au segment (a) par la relation *Élaboration d'entité* et les segments (b) et (c) seraient considérés comme élaborant l'entité *raisons*. Cette solution n'est pas convaincante parce qu'ainsi la relation d'*Explication* à l'œuvre disparaît totalement de la représentation.

La deuxième solution accorde à l'amorce un rôle discursif primordial avec le traitement de celle-ci comme un constituant du discours.

Solution 2 : Établir une relation de discours au sein même de l'amorce

Cette solution vise à considérer le segment (a) de (21) non pas comme un segment unique mais comme deux segments de la façon suivante :

=(21) [Domenech refuse obstinément cette concession] $_{a_1}$ [pour **deux raisons**] $_{a_2}$.
 [D’abord, il ne la comprend pas.] $_b$ [Ensuite, en l’acceptant, il aurait le sentiment de ruiner
 une autorité déjà amplement chancelante.] $_c$

Le segment a_2 est alors attaché au segment a_1 par la relation d’*Explication*. Le segment a_2 est quant à lui élaboré par les segments (b) et (c), tel que la figure 6.7 le représente :

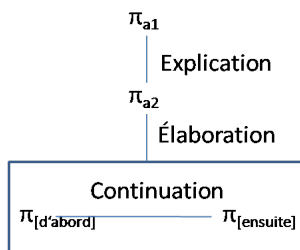


FIG. 6.7 – Structure de (21)

Cette solution, au premier abord, soulève des questions quant à la nature des segments discursifs. On considère généralement qu’une unité de discours élémentaire correspond à une proposition syntaxique. Or, nous n’avons qu’une proposition contenue dans le segment (a).

Les configurations des exemples suivants constituent des arguments empiriques en faveur de cette solution :

- (27) [Ces fossés ont, qui plus est, tendance à s’agrandir.], $_a$ [et ce pour deux raisons :] $_b$
 [d’abord, les accords de branche ou les accords d’entreprise qui se multiplient à
 l’intérieur des frontières nationales créent davantage de situations particulières ;] $_c$
 [ensuite, l’individualisation du temps de travail risque, comme le note l’enquête,
 de «porter le dernier coup à des modèles nationaux déjà bien malmenés»]. $_d$ (Le
 Monde)
- (28) [Pour certains optimistes, il n’y a pas lieu de s’inquiéter d’une telle évolution.], $_a$
 [Et ce pour trois raisons.], $_b$ [Premièrement, la croissance actuelle de l’écono-
 mie française, sur un rythme de 4 % l’an, démontrerait que des flux de création
 d’entreprises plus faibles ne remettent pas nécessairement en cause la prospérité
 nationale (...).] $_c$ [Deuxièmement, les défaillances d’entreprises sont également à un
 faible niveau, le plus faible depuis 1989 (...).] $_d$ [Troisièmement, la question de la
 pérennité des entreprises créées devrait être prise en compte pour mesurer plus
 exactement la réalité de la situation.]. $_e$ (Le Monde)
- (29) [La synthèse chimique ouvre ce que Philippe Amade appelle une «voie royale»]. $_a$
 [Pour deux raisons.], $_b$ [D’abord, l’essor de la chimie combinatoire permettra de
 créer de nouvelles molécules synthétiques à partir de composés marins.]. $_c$ [Ensuite
 - et surtout - grâce à la microbiologie, on pourra cultiver le micro-organisme qui
 produit la molécule recherchée.]. $_d$ (Le Monde)

Dans les exemples ci-dessus, l’amorce constitue à elle toute seule un segment de discours. En (27), celle-ci est détachée et introduite par le pronom démonstratif indéfini «ce» qui renvoie à la proposition décrite dans le segment (a). En (28) et (29), c’est une phrase

ponctuationnelle. Chacune de ces trois configurations est une paraphrase possible pour (21).

Ces données et manipulations montrent l'autonomie de l'amorce. Et pour ces cas, le segment (b) est donc traité comme un constituant, attaché au segment (a) par *Explication* et élaboré par les segments suivants.

Par opposition à la première solution, cette dernière favorise l'idée que les items de l'énumération élaborent l'amorce tout en mettant au second plan les relations s'établissant entre chacun des items et la phrase amorce.

L'amorce joue le rôle d'un topique discursif dominant les items de l'énumération et donnant des indications sur la relation de discours à l'oeuvre entre l'amorce et les items à travers le critère de co-énumérabilité annoncé par l'énumérathème, ici les items sont classifiés comme des «raisons». L'introduction de ce topique de discours dans la structure soulève la question de la nécessité de rendre compte de l'amorce au moyen d'un constituant. Bien que ce constituant semble jouer un rôle de facilitation dans l'inférence de la relation de discours entre les items de l'énumération et l'amorce, il faut justifier l'insertion de ce constituant afin d'éviter une prolifération non nécessaire de ces topiques dans la structure. Une façon de justifier la présence de ce topique est de vérifier si ce dernier est disponible pour la poursuite du discours, comme en atteste le segment (f) dans l'exemple suivant :

- (30) [Pour certains optimistes, il n'y a pas lieu de s'inquiéter d'une telle évolution.]_a
[Et ce pour trois raisons.]_b [Premièrement, la croissance actuelle de l'économie française, sur un rythme de 4 % l'an, démontrerait que des flux de création d'entreprises plus faibles ne remettent pas nécessairement en cause la prospérité nationale (...).]_c [Deuxièmement, les défaillances d'entreprises sont également à un faible niveau, le plus faible depuis 1989 (...).]_d [Troisièmement, la question de la pérennité des entreprises créées devrait être prise en compte pour mesurer plus exactement la réalité de la situation.]_e **[Ces arguments, apparemment séduisants, sont malheureusement contestables.]**_f

Le segment (f) constitue la clôture de la structure énumérative. Cette clôture est composée, comme l'amorce, d'un énumérathème, ici «arguments», qui condense en une expression référentielle les items énumérés. Les clôtures ont pour rôle de conclure la structure énumérative (Ho-Dac *et al.*, 2009). Le meilleur site d'attachement pour le segment (f) est de notre point de vue le segment (b), topique discursif de l'énumération dans lequel l'ensemble des items est annoncé et introduit dans le constituant. Le segment (f) est, de notre point de vue, attaché au segment (b) par la relation de *Contraste*.

Le rôle de la clôture étant de conclure la structure énumérative, généralement en reprenant l'ensemble des items, il apparaît donc qu'un constituant contenant l'ensemble des items doit être disponible dans la structure. Le constituant contenant l'amorce va jouer ce rôle.

Cependant, la création de ce topique de discours n'apparaît pas toujours évidente. Dans les cas typiques d'amorce, l'énumérathème ne peut être scindé de la phrase amorce, comme en (31) :

- (31) [L'hygiène dentaire (...) se fonde sur plusieurs éléments.]_a [**D'abord**, il importe d'assurer un bon nettoyage des dents.]_b [Puis il faut savoir faire appel au dentiste dès que l'on constate quelque chose d'anormal.]_c

Pour rendre compte du rôle primordial de l'amorce et des marqueurs d'items, que l'amorce puisse ou non faire l'objet d'un constituant du discours, nous avons introduit, dans (Bras *et al.*, 2008), une troisième solution consistant à ajouter une nouvelle relation de discours dans le cadre de la SDRT, la relation d'*Énumération*.

Solution 3 : Introduire une nouvelle relation de discours *Énumération*

Cette solution vise à introduire dans le cadre de la SDRT une nouvelle relation de discours, *Énumération*, permettant de rendre compte plus précisément du type d'information disponible à la surface du texte (l'amorce et les marqueurs d'item) constitutifs de la structure énumérative. *Énumération* est une relation subordonnante, classée parmi les relations structurelles au même titre que *Parallèle* et *Contraste*, en opposition aux relations de contenu comme *Élaboration* ou *Explication*.

L'inférence de cette relation structurelle rend compte d'informations textuelles. Elle peut, à première vue, paraître inutile ou redondante dans les cas où une relation de contenu subordonnante, par exemple *Élaboration*, est inférable à partir du contenu des segments de la structure énumérative.

Introduire cette relation de discours revient à considérer le rôle des informations apportées par la structure énumérative sur l'interprétation globale du discours. Pour illustrer ce point, Bras (2007) analyse l'exemple suivant :

- (32) =(22) [Tout en roulant vers l'île Saint-Louis, il prit deux décisions.]_a [**D'abord**, il allait persuader sa mère de quitter Paris au moins jusqu'au procès et de descendre dans le Midi, peut-être même en Italie où elle ne risquerait rien.]_b [Et puis, de son côté, il abandonnerait la rue de Longchamp et louerait un studio meublé.]_c (Michel Droit, *Le Retour*, 1964, p. 298)

Il y a, de notre point de vue, deux façons d'appréhender les processus interprétatifs à l'œuvre à la lecture de cette structure énumérative : soit on considère que seules des connaissances du monde interviennent dans le fait d'interpréter que «persuader quelqu'un» ou «abandonner un appartement» sont des événements qui peuvent être des décisions ; soit on considère que l'ensemble de la structure énumérative et le prospect, composé du numéral «deux» annonçant le nombre d'items de la structure et de l'énumérathème «décisions» spécifiant le critère de co-énumérabilité, obligent à inférer que les événements décrits en (b) et (c) sont des décisions.

Les informations, se situant au niveau des types des éventualités, encodées dans le prédicat *Subtype_D*, sont en théorie disponibles mais, en pratique, difficiles à établir. En revanche, les informations de surface (amorce, marqueurs d'item) constitutifs de la structure énumérative sont mises à disposition et facilitent, voire orientent, l'interprétation. L'idée de Bras *et al.* (2008) est donc d'exploiter ces informations pour l'inférence de la relation d'*Énumération*, même si, dans certains cas, d'autres relations de contenu pourraient être inférées au plan théorique, rendant inutiles l'inférence de cette dernière.

De plus, en (22), le prédicat *Subtype_D* est vérifié entre un type placé très haut dans la hiérarchie, donc fortement sous-spécifié, et «persuader quelqu'un» et «abandonner un appartement». Bras *et al.* (2008) soulèvent la possibilité d'exclure *Élaboration*, relation qui n'autorise pas la mise à jour du topique, pour traiter ce genre de cas, comme en (31).

- (33) = (31) [L'hygiène dentaire (...) se fonde sur plusieurs éléments.]_a [D'abord, il importe d'assurer un bon nettoyage des dents.]_b [Puis il faut savoir faire appel au dentiste dès que l'on constate quelque chose d'anormal.]_c

En (22) et (31), les amorces contiennent un énumérathème qui a pour caractéristique d'être un nom abstrait sous-spécifié en attente de spécifications. L'amorce introduit donc un constituant explicite en attente de spécifications qui doit être mis à jour au fur et à mesure de l'interprétation de la structure. Inférer la relation de *Topique*, contrairement à la relation d'*Élaboration*, permet cette mise à jour.

Une structure énumérative pose des contraintes, par exemple le nombre d'items annoncé dans l'amorce doit être satisfait dans l'énumération pour garantir une cohérence maximale. Une formalisation complète de la relation d'*Énumération* permettrait de contrôler que la structure énumérative est bien formée en vérifiant que le nombre d'items annoncé dans l'amorce est atteint et non dépassé et également de contrôler que les marqueurs de premier item tels que *d'abord*, *dans un premier temps*, de suite tels que *ensuite* ou de dernier item tels que *enfin* sont correctement ordonnés. Bras (2007) suggère également que déterminer la nature temporelle ou non temporelle de l'amorce et de l'énumération qu'elle introduit permettrait de désambiguïser des interprétations où *puis* est temporel de celles où il est purement énumératif.

L'inférence de la relation d'*Énumération* s'appuie sur la reconnaissance d'une amorce et de marqueurs d'items. Une première formalisation de la relation d'*Énumération* est proposée dans (Bras, 2007), puis une généralisation est proposée dans (Bras *et al.*, 2008). Les rôles respectifs de *d'abord*, *dans un premier temps* et *premièrement* ont été décrits dans la section 6.1. C'est uniquement leur rôle de marqueurs de premier item (ci-après MPI) qui est, ici, considéré.

Axiome 6.2 Inférer *subord* *coord* *MPI*

$$(?(\alpha, \beta, \lambda) \wedge MPI(\beta)) > (\Downarrow (\alpha, \beta, \lambda) \wedge \exists \gamma \exists R \in Coord R(\beta, \gamma, \lambda)) \wedge item(\gamma)$$

Cet axiome permet d'inférer une relation subordonnante sous-spécifiée en présence d'un marqueur de premier item et de préparer l'interprétation d'au moins un deuxième segment (γ dans l'axiome), qui sera attaché à celui contenant le marqueur de premier item par une relation coordonnante.

La présence conjointe d'une amorce et d'un marqueur de premier item permet l'inférence de la relation d'*Énumération* :

Axiome 6.3 Inférer *Énumération*

$$(?(\alpha, \beta, \lambda) \wedge amorce(\alpha) \wedge MPI(\beta)) > \dot{E}numeration(\alpha, \beta, \lambda)$$

La relation d'*Énumération* est une relation subordonnante. Elle implique $\Downarrow (\alpha, \beta, \lambda)$, avec mise à jour du topique.

Enfin, Bras *et al.* (2008) ajoutent à ces deux axiomes, un axiome qui rend compte du rôle des marqueurs d'autres items (ci-après MAI) de suite ou de dernier item pouvant suivre un MPI.

Axiome 6.4 *Inférer_subord_coord_MAI*

$(?(\beta, \gamma, \lambda) \wedge \exists \alpha \text{ Énumération}(\alpha, \beta, \lambda) \wedge \text{MAI}(\gamma)) > (\text{Énumération}(\alpha, \gamma, \lambda) \wedge \text{item}(\gamma) \wedge \text{Continuation}(\beta, \gamma, \lambda))$

Énumération est inférée entre l'amorce et chaque item, permettant d'établir, selon le principe inverse de la distributivité, la relation coordonnante *Continuation* entre les constituants représentant les items. Cet axiome pourrait être complété par des axiomes spécifiques aux marqueurs de suite comme *puis* ou *ensuite* pouvant introduire une relation coordonnante plus riche comme *Narration* dans les cas où l'amorce est temporelle ou événementielle.

Tandis que les MPI sont traités comme s'attachant au contexte gauche par une relation subordonnante et au contexte droit par une relation coordonnante, les MAI sont traités comme s'attachant par deux fois au contexte gauche, à l'amorce au moyen d'une relation subordonnante *Énumération* et à l'item le précédant directement par une relation coordonnante *Continuation*. Ce dernier axiome permet de traiter les marqueurs de série tels que *puis*, *ensuite*, *etc.* et les marqueurs de dernier item tels que *enfin* ou les ordinaux précédés de *et*, comme *et troisièmement*, *etc.* Un axiome pourrait déterminer le rôle plus précis des marqueurs de dernier item en s'assurant qu'un autre item ne suive pas celui introduit par ce dernier.

La relation d'*Énumération* que nous proposons permet de construire des structures qui peuvent être rapprochées des structures N-aires de Polanyi *et al.* (2004) si l'on considère que le contenu des items et également la sémantique de la relation coordonnante qui lie les items soient mis à jour dans le topique sous-spécifié. Ce point est appuyé par des structures énumératives pour lesquels l'amorce signale la relation s'établissant entre les items, comme ci-dessous :

- (34) (...) les Européens sont divisés vis-à-vis de la Yougoslavie, à l'OTAN, comme au GATT ; sur la menace japonaise comme sur toutes les questions de politique économique, [deux attitudes opposées s'affrontent :]_a [celle des pays du Nord, désireux de créer une zone de libre-échange dépendant des Etats-Unis, et, surtout, soucieux d'en faire le moins possible au plan européen,]_b [et celle, plus latine, des Etats industriels désireux de construire ensemble leur avenir.]_c

La relation de *Contraste* qui lie les items (b) et (c) est ici signalée dans l'amorce (a) «deux attitudes opposées s'affrontent».

Solution 4 : Tirer parti d'une combinaison d'indices

Enfin, nous proposons, dans cette thèse, une quatrième solution qui est un prolongement de la solution 3 présentées ci-dessus.

Cette proposition consiste à tirer parti des informations contenues dans l'amorce une fois la relation d'*Énumération* inférée. Le SN prospectif est alors traité comme un connecteur

signalant la relation de discours à l'œuvre entre l'amorce et l'énumération ou entre les items de l'énumération.

En (6), le SN prospectif «en deux temps» est marqueur d'une *Élaboration* s'établissant entre l'amorce et l'énumération. En (21), le SN prospectif «pour deux raisons» est marqueur d'une *Explication*. En (34), le SN prospectif «deux attitudes opposées» est marqueur d'un *Contraste* entre les items de l'énumération. Cette solution revient à considérer la présence d'une structure énumérative comme déclencheur d'une relation de discours de contenu.

On peut inférer la relation d'*Élaboration* lorsque le SN prospectif contient un énumérathème qui est un nom d'événement ou de temps ($\text{Énumérathème}_{\text{événement}}$ dans l'axiome). On propose donc la règle d'inférence suivante :

Axiome 6.5 *Inférer_Énumération_Élaboration*

$$(?(\alpha, \beta, \lambda) \wedge \text{Énumération}(\alpha, \beta, \lambda) \wedge \text{Énumérathème}_{\text{événement}}[\alpha] > \text{Élaboration}(\alpha, \gamma, \lambda))$$

Nous proposons une règle défaisable dans les cas où un autre marqueur tel que *parce que* signalant une relation d'*Explication* est présent. L'inférence de la relation d'*Explication* sera, ici, prioritaire sur celle d'*Élaboration*.

Pour continuer la formalisation des structures énumératives, nous allons considérer des cas d'items structurés et montrer qu'il est nécessaire de traiter les marqueurs d'item comme des introducteurs de nouveau topique.

6.2.2.2 Marqueurs d'item comme introducteurs de cadre de discours

- (35) [Ces exemples appellent à la prudence.]_a [Les facteurs essentiels sont ici les suivants.]_b
[En premier lieu,]_c [la guerre a en général pour les acteurs locaux un intérêt politique de premier ordre, souvent bien plus important que celui des Etats- Unis.]_{c1}
 [Leur tolérance à la souffrance est donc plus grande.]_{c2} **[En deuxième lieu,]** en dépit de leur taille réduite, ces acteurs supplantent d'ordinaire les Etats-Unis dans une ressource précise : le nombre d'hommes en âge de combattre. Même s'il n'est plus l'élément déterminant de la guerre terrestre, il reste un facteur critique, notamment en ville, dans la jungle ou en montagne.]_d **[Troisièmement,]** les «locaux» disposent en général d'un avantage (...).] **[Quatrièmement,]** nombre des chefs militaires de ces Etats ou entités ont été formés dans le monde développé - (...).] **[Cinquièmement,]** l'arsenal nécessaire au combat rapproché, (...) est beaucoup moins coûteux que les armements nécessaires à la guerre dans les "espaces communs".]_g

Comme nous pouvons le voir en (35), les items peuvent être des constituants discursifs complexes, i.e. constitué de plusieurs constituants simples, par exemple le premier item introduit par *en premier lieu* est composé de deux constituants simples obtenus à partir des segments (c₁) et (c₂). Le marqueur d'item *en premier lieu* a ici une portée au-delà de la phrase syntaxique dans laquelle il se trouve. Schnedecker (2001b) et Jackiewicz (2002), entre autres, ont montré que les marqueurs d'intégration linéaire en position détachée à l'initiale ont un rôle cadratif suivant les hypothèses de Charolles (1997), c'est-à-dire qu'ils introduisent un cadre pouvant aller au-delà de la phrase.

Suivant la formalisation de Vieu *et al.* (2005) pour les adverbiaux de localisation spatio-temporelle détachés à l'initiale, nous proposons de traiter les marqueurs d'item détachés à l'initiale comme des introducteurs de nouveaux topiques sous-spécifiés, considérés comme des constituants du discours à part entière, qui dominent discursivement les segments dans la portée du cadre, i.e. les segments constitutifs de l'item. En (35), le marqueur d'item *en premier lieu* permet de grouper les constituants π_{c_1} et π_{c_2} sous la dominance de π_c dont le contenu est équivalent à «ce qui est à prendre en compte en premier lieu».

Le rôle discursif de ces marqueurs d'item détachés à l'initiale est non seulement justifié au niveau linguistique mais également du point de vue de la structuration du discours. Comme le montre l'exemple (36) (extrait de (35)), si l'on ne tient pas compte du rôle du marqueur d'item, l'attachement de π_{c_2} à π_{c_1} est problématique :

- (36) [Les facteurs essentiels sont ici les suivants.]_b [En premier lieu, la guerre a en général pour les acteurs locaux un intérêt politique de premier ordre, souvent bien plus important que celui des Etats- Unis.]_{c_1} [Leur tolérance à la souffrance est donc plus grande.]_{c_2}

En (36), le constituant π_{c_1} dont le segment est introduit par *en premier lieu* est attaché à π_b par *Énumération*. π_{c_1} , en tant que premier item de l'énumération, est le premier argument d'une relation coordonnante introduite, non pas par l'item suivant mais par une proposition incluse dans l'item en cours de développement : π_{c_2} . Cet attachement a pour conséquence que le constituant π_{c_1} ne se trouve plus sur la frontière droite et n'est donc plus disponible pour l'attachement de l'item suivant décrit dans le segment (d). L'attachement du constituant π_d devient donc fortement problématique, Cf. Figure 6.8.

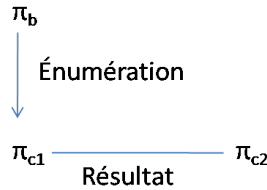


FIG. 6.8 – Structure de (36)

L'introduction d'un niveau hiérarchique supplémentaire permet d'éviter ce problème, Cf. Figure 6.9.

L'information apportée par les marqueurs d'item concerne uniquement la structuration de l'énumération. Ils n'ont pas de contenu sémantique, comme le souligne Schnedecker (2001b). Elle oppose ainsi les cadratifs spatio-temporels et les marqueurs d'intégration linéaire. Seuls ces derniers peuvent être supprimés car ils n'ont justement pas de contenu sémantique. Le traitement proposé pour les marqueurs d'item est le suivant :

$$\exists \pi, \pi : [\exists P \exists e (P(e) \wedge \text{Position} - \text{enum}(e))]$$

Ce traitement spécifie qu'au marqueur d'item correspond un constituant π fortement sous-spécifié qui précise seulement qu'une éventualité e va être introduite et donne sa position dans l'énumération.

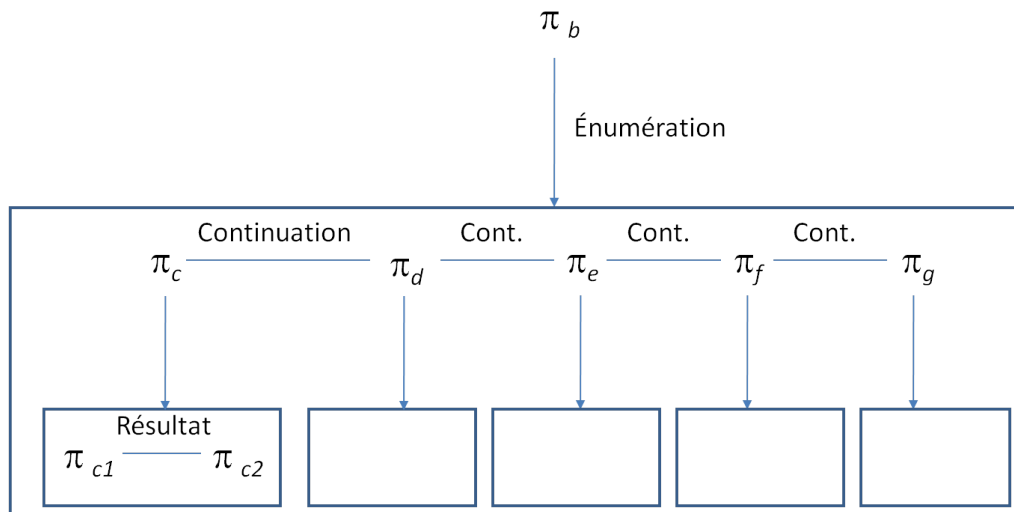


FIG. 6.9 – Structure de (35)

L'introduction d'un niveau hiérarchique supplémentaire présente aussi l'avantage de procurer une structure particulièrement homogène, comme illustré en Figure 6.9, qui pourra faciliter, entre autres, la détection d'énumérations mal formées et permet de rendre compte du rôle facilitateur des adverbiaux détachés à l'initiale qui favorise, au niveau visuel, la création de blocs textuels.

6.2.2.3 *Élaboration* des topiques

L'étude des structures énumératives a permis d'alimenter le débat sur le statut des topiques discursifs en SDRT. Les topiques de discours peuvent s'organiser sur une échelle allant du topique explicite au topique implicite.

Les constituants explicites correspondent à des segments de discours qui sont reliés à leurs éléments subordonnés par la relation d'*Élaboration*. A l'autre extrémité de l'échelle, les constituants implicites sont inférés grâce à la présence de plusieurs segments de discours coordonnés qui partagent un topique commun reconstituable à partir du contenu de chacun des segments subordonnés et sont liés par la relation de *Topique* ou \Downarrow .

Les amorces et les marqueurs d'item cadratifs forment des topiques fortement sous-spécifiés et font intervenir un type de topique qui n'est ni explicite, ni implicite. Nous qualifions ces topiques de semi-implicites. Ils sont associés à des formes linguistiques (et ne sont donc pas implicites). En revanche ils doivent être mis à jour (contrairement aux constituants explicites).

La division effectuée par le passé en SDRT entre les relations d'*Élaboration* (entre un constituant **explicite** et des constituants subordonnés) et de *Topique* (entre un constituant **implicite** et des constituants subordonnés) est réductrice par rapport à la richesse des objets décrits ici. La relation d'*Énumération* pourrait servir d'intermédiaire entre les deux. Celle-ci mettrait en jeu un constituant explicite fortement sous-spécifié et la mise à jour de

ce constituant avec le contenu des items et la relation qui les relie, suivant les propositions faites par Polanyi *et al.* (2004) pour les structures n-aires.

Il faudrait à présent pouvoir rendre compte du degré de spécification de l’amorce qui influe sur la possibilité d’inférer *Élaboration* (dans le cas où l’amorce est suffisamment spécifiée) ou *Énumération* (dans la cas où l’amorce est fortement sous-spécifiée) ou encore de conclure à un discours incohérent (en cas d’amorce sous-spécifiée incompatible avec les items).

L’énumérathème ou classificateur énumératif a été décrit comme imposant une contrainte d’identité sémantique sur les items (Schnedeker, 2001a). La plasticité de cette contrainte lors de l’interprétation n’est pas clairement définie. En (37), le constituant construit pour l’amorce reçoit un contenu sémantique fortement sous-spécifié et va être mis à jour par le contenu des items qui devront dans la mesure du possible vérifier les contraintes posées par l’énumérathème :

- (37) Hier, j’ai pris deux décisions : premièrement qu’il pleuvrait ce matin, deuxièmement que la lune serait pleine à la fin de la semaine.

Il s’ensuit, en (37), une étrangeté liée à la tension entre l’énumérathème et le contenu des items. Certains locuteurs vont chercher un contexte qui permette de satisfaire cette tension, se basant sur le fait que nous avons là une structure énumérative bien formée. Les autres vont le juger incohérent. En SDRT, nous pouvons nous appuyer sur la nouvelle relation d’*Énumération* et montrer que cette structure énumérative est bien formée au niveau structurel. Par ailleurs nous n’inférons pas de relation de contenu.

6.2.3 Synthèse

Cette étude a permis de montrer que toutes les structures énumératives ne contiennent pas des élaborations. Ce sont cependant des structures intéressantes pour la recherche d’*Élaboration* dans notre corpus. Nous avons par ailleurs mis au jour un marqueur de la relation mettant en jeu plusieurs indices pointant vers l’inférence de celles-ci.

La distinction entre *Élaboration* et *Topique* est basée sur des critères théoriques, i.e. sur la mise à jour ou non des topiques. En ce qui nous concerne, ces deux relations peuvent être confondues. Elles sont par ailleurs signalées par les mêmes indices.

Jusqu’à maintenant nous n’avons examiné que des exemples prototypiques pour lesquels les structures textuelles et sémantiques sont isomorphes. Il faut maintenant s’attaquer à des cas moins canoniques comme ceux étudiés par Luc (2000) dans lesquels la structure textuelle semble orthogonale à la structure sémantique. Pour vérifier que les structures énumératives peuvent bien être hétéromorphes, il faut être capable de séparer le niveau textuel du niveau sémantique et comparer leur contribution de manière systématique.

En s’appuyant sur notre proposition de traitement des structures énumératives, nous allons considérer les structures énumératives à deux temps, qui vont, comme nous allons le voir, constituer un défi pour le traitement en SDRT du fait que l’attachement du second temps semble traverser la frontière droite.

6.3 Structures énumératives à deux temps

Les structures énumératives à deux temps ont fait l'objet d'une étude au cours de notre thèse (Vergez-Couret *et al.*, 2008) dans laquelle nous avons proposé un traitement fidèle au principe de la frontière droite.

6.3.1 Description des structures énumératives à deux temps

Les structures énumératives classiques, que nous avons illustrées dans les sections précédentes, sont composées d'une amorce (le segment (a) dans l'exemple (38)) et d'une série d'items (deux en (38) (b-d) et (e-g)). Le second temps des structures énumératives dites à deux temps (Porhiel, 2007), suivant l'amorce et la première énumération, est illustré par les segments (h-l) et (m-o).

- (38) [Deux fois, Nil crut sombrer dans le gouffre :]_a [**d'abord** chez Vrin,]_b [où il feuilletait des livres avec Anne-Geneviève,]_c [quand Laure y entra inopinément ;]_d [**ensuite** à la bibliothèque de Beaubourg,]_e [où il était assis en face de Laure,]_f [et où Anne-Geneviève apparut soudain.]_g [Chez Vrin,]_h [dès qu'il avait vu Laure,]_i [Nil s'était lâchement jeté dans l'arrière-boutique,]_j [pour n'en sortir]_k [qu'après s'être assuré que son amante numéro deux ne se trouvait plus dans la librairie.]_l [À Beaubourg,]_m [Anne-Geneviève n'était, Dieu merci, restée qu'un instant (...)]_n, [et Laure n'avait même pas levé le nez de ses cahiers.]_o (Gabriel Matzneff, *Ivre du vin perdu*, 1981)

L'amorce se caractérise par la combinaison d'un SN prospectif «deux fois» et d'un marqueur typographique (:). Les items sont introduits par la combinaison d'un marqueur d'item, respectivement «d'abord» et «ensuite», et d'un introducteur de cadre, respectivement «chez Vrin» et «à Beaubourg». Suivant Vieu *et al.* (2005), ces cadres introduisent des topiques discursifs dominant les phrases sémantiquement sous leur portée. Ici, le segment (b) domine (c-d) et le segment (e) domine (f-g). Le second temps est construit en faisant un rappel des items de la première énumération en répétant les introducteurs de cadre, «Chez Vrin» en (h) et «À Beaubourg» en (m).

Des exemples similaires ont été étudiés par Schnedecker (2007) dans le contexte d'une étude sur les corrélats anaphoriques, tels que *le premier*, *le second*. Schnedecker (2004) parle de discours à double ou à multi-topique. Les discours à double-topique, auxquels nous nous limiterons ici, sont organisés autour de la description de deux référents. Les expressions permettant d'y référer ont la forme de couples de pronoms anaphoriques, tels que *l'un/l'autre* ou *le premier/le second*. Ces discours sont caractérisés par des relations de parallélisme entre les structures syntaxiques et sémantiques, comme l'illustre l'exemple (39).

- (39) [Les organisations de salariés doivent choisir entre deux options stratégiques :]_a [soit se laisser intégrer dans les pactes nationaux pour la compétitivité,]_b [soit tenter de s'en dégager.]_c [Dans le premier cas,]_d [elles ne défendront les salariés que dans la mesure où cette compétitivité de l'économie nationale le permettra (...)]_e [Dans le second cas,]_f [elles peuvent tenter de retrouver leur capacité de mobilisation et se reconstituer en tant que force d'opposition.]_g (Schnedecker, 2007, p. 307)

Le lien anaphorique entre le premier temps (b-c) et le second temps (d-g) est, ici, supporté par un mécanisme de référence complexe faisant intervenir les corrélats anaphoriques *le premier* et *le second*.

Pour analyser ces structures croisées au moyen de relations de discours, un traitement naïf de ces deux exemples se heurte à un problème dès que l'attachement du second temps est envisagé. En (38), suivant les propositions faites dans les sections précédentes, le traitement de la première étape mène à une structure hiérarchique dans laquelle l'amorce domine les items de l'énumération. L'amorce décrivant un événement qui se décompose en deux temps «deux fois», nous pouvons inférer, en plus d'*Énumération*, la relation d'*Élaboration*. Les items sont situés au même niveau de l'organisation discursive, reliés par *Parallèle*. De ce fait, selon le principe de la frontière droite (décrit page 19), les deux items ne peuvent pas se trouver simultanément sur la frontière droite. Seulement le dernier item introduit (e) peut s'y trouver, *Cf.* Figure 6.10.

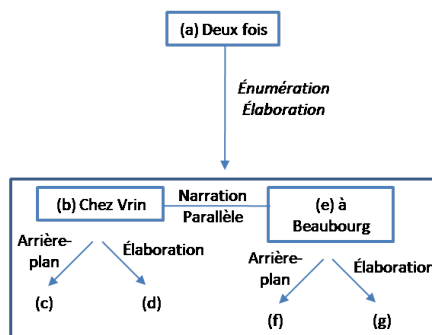


FIG. 6.10 – Représentation de la première étape de (38)

Cependant, dans le cas des structures énumératives à deux temps, les deux items du second temps ont été décrits comme disponibles pour des développements futurs.

En (38), deux histoires sont croisées, racontant comment, par deux fois, Nil eut peur de sombrer dans le gouffre. Dans l'histoire se déroulant chez Vrin, Nil est avec Anne-Geneviève lorsque Laure apparaît dans le premier temps. Dans le second temps, dès que Nil aperçoit Laure, celui-ci s'échappe dans l'arrière-boutique. Le premier item du premier temps et celui du second temps forment une séquence temporelle. Dans l'histoire se déroulant à Beaubourg, Nil est avec Laure lorsqu'Anne-Geneviève apparaît dans le premier temps. Dans le second temps, il est décrit que les deux amantes ne se voient pas. Le deuxième item du premier temps et celui du second temps forment également une séquence temporelle. On relève donc deux structures : l'une affleurant à la surface du texte et l'autre relevant d'une interprétation sémantique plus profonde.

Ces structures relèvent des *structures croisées* (*interleaved structures*) analysées dans plusieurs théories du discours (Asher, 1993; Wolf et Gibson, 2005; Egg et Redeker, 2008; Danlos, 2008; Danlos et Hankach, 2008).

6.3.2 Des structures similaires dans les théories du discours

Pour traiter de telles structures en SDRT, Asher (1993) propose le principe de la «subordination discursive» (*discourse subordination mechanism*). Selon ce principe, la construction de la représentation s'organise autour des topiques introduits. Dans ce traitement, les noms propres sont supposés capables de traverser la frontière droite afin de mettre à jour le constituant fermé avec des informations additionnelles. Il illustre ce principe avec l'exemple suivant :

- (40) There are two houses you might be interested in : House A is in Palo Alto, House B in Portola Valley. Both were built in 1950, and both have 3 bedrooms. House A has 2 baths and B, 4. House A also has a kidney shaped pool. House A is on quarter acre, with a lovely garden, while House B is on 4 acres of steep wooded slope, with a view of the mountains. The owner of House B is asking \$425K. The owner of house A is asking \$600K. (Asher, 1993, p. 345)

Pour justifier un tel traitement, Asher (1993) suggère qu'il existe deux référents du discours disponibles pour parler des informations concernant chaque maison, comme la poursuite en (41) le souligne :

- (41) That's all I know about House A, but I will know more about House B tomorrow.

Cette solution prend les objets eux-même (ici les maisons) comme principe organisateur, les considérant comme des topiques discursifs guidant la construction du discours, Cf. Figure 6.11.

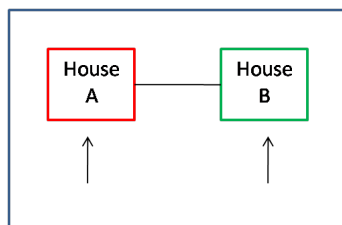


FIG. 6.11 – Les «houses» comme principe organisateur

Cependant ce traitement entraîne un mécanisme complexe qui viole le principe de la frontière droite. L'ordre dans lequel les constituants sont introduits est complètement perdu. Ce traitement n'est de plus validé par aucun traitement discursif, par exemple la résolution anaphorique. Enfin, ce principe introduit dans (Asher, 1993) n'est pas repris dans (Asher et Lascarides, 2003) laissant la question du traitement des structures croisées ouverte. Nous ne pensons pas toutefois pouvoir proposer, à ce stade, un traitement homogène pour ces exemples tels que (40) et les structures énumératives à deux temps telles que (38) et (39).

D'autres approches du discours ont également fait des propositions pour le traitement des structures croisées. Nous allons comparer ci-dessous les propositions de Wolf et Gibson

(2005), Egg et Redeker (2008) et Danlos (2008).

Les analyses de Wolf et Gibson (2005) cherchent à rendre compte de toutes les relations s'établissant dans le discours, les relations données par le niveau surfacique aussi bien que celles obtenues au niveau de la structure sémantique profonde. La représentation de (38) dans leur cadre ne tient pas compte des introducteurs de cadre («À Beaubourg», «Chez Vrin»...) comme introduisant des constituants discursifs. Ces derniers sont donc inclus dans le segment ultérieur ((b) est inclus dans (c) ; (e) est inclus dans (f) ; (h) est inclus dans (i) et (m) est inclus dans (n)). Dans leur ensemble de relations de discours, Wolf et Gibson (2005) distinguent les relations asymétriques, comme *Élaboration* ou *Séquence temporelle* (*Narration* en SDRT) des relations symétriques comme *Contraste* ou *Similarité* (*Parallèle* en SDRT). Cette distinction est proche des notions de noyau et satellite de la RST. Les relations asymétriques se tiennent entre un noyau et un satellite tandis que les relations symétriques relient deux noyaux. Cette distinction semble légèrement différente de celle des relations subordonnantes et coordonnantes (voir (Busquets *et al.*, 2001)). Les relations subordonnantes et coordonnantes imposent des contraintes sur la structure, suivant le principe de la frontière droite, tandis que les relations symétriques et asymétriques ne posent aucune contrainte. La figure 6.12 illustre les dépendances croisées (*crossed dependencies* (Wolf et Gibson, 2005)) sur l'exemple (38) non autorisées en SDRT.

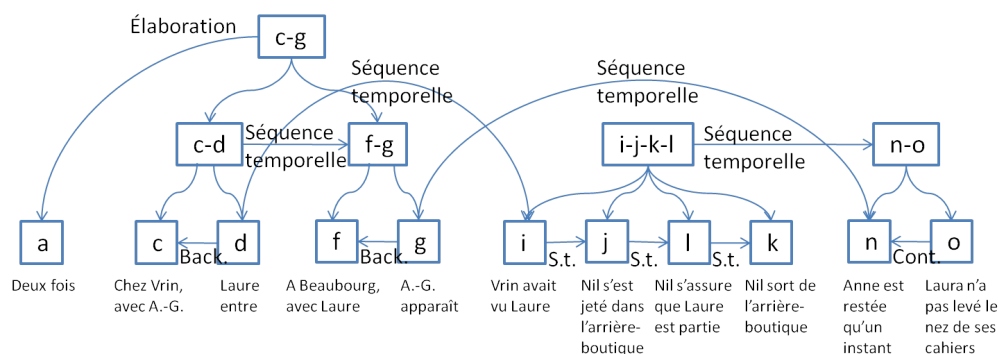


FIG. 6.12 – Les dépendances croisées (Wolf et Gibson, 2005) illustrées avec l'exemple (38)

Les dépendances croisées apparaissent entre les deux temps de la structure énumérative, plus précisément une relation de *Séquence temporelle* est établie entre (d) (premier item du premier temps) et (i) (premier item du second temps), et également entre (g) (deuxième item du premier temps) et (n) (deuxième item du second temps).

Egg et Redeker (2008) rejettent l'analyse de Wolf et Gibson (2005), pointant que celle-ci ne présente pas les contraintes apportées par les informations de surface. L'analyse de Egg et Redeker (2008) donne la priorité aux marqueurs de surface. En (38), ces marqueurs de surface sont donnés par les introducteurs de cadre, qui renforcent la cohérence et signalent la relation de *Parallèle*. Egg et Redeker (2008) donnent priorité au niveau surfacique en ne tenant pas compte du niveau sémantique, Cf. Figure 6.13.

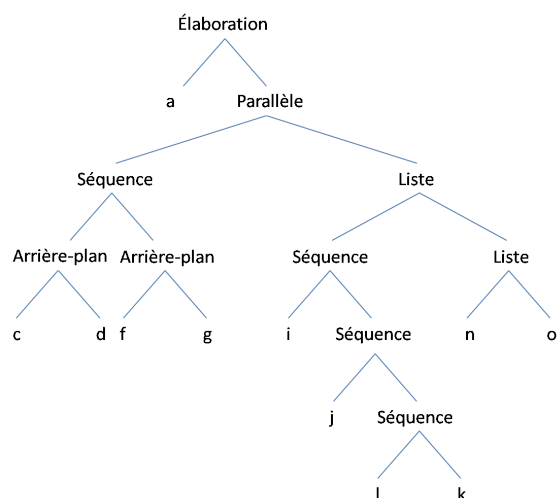


FIG. 6.13 – Représentation de (38) chez Egg et Redeker (2008)

Enfin, Danlos et Hankach (2008) sont en accord avec l'analyse de Egg et Redeker (2008). Ils ne représentent pas les liens anaphoriques entre les introducteurs de cadre et, selon eux, ceux-ci ne peuvent pas être systématiquement représentés dans la structure du discours, Cf. Figure 6.14.

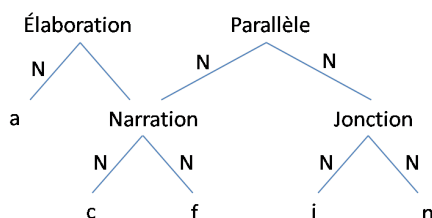


FIG. 6.14 – Représentation de (38) chez Danlos (2008)

Les deux niveaux, niveau textuel et niveau sémantique, pointent des structures différentes. Le niveau textuel présente deux énumérations parallèles (respectant le principe de la frontière droite) tandis que le niveau sémantique présente une structure temporelle (qui traverse la frontière droite). L'accessibilité anaphorique au moyen d'outils de référence spécifiques tels que les descriptions définies, les corrélats anaphoriques, comme en (39), ou la répétition d'introducteur de cadre en (38), brouille la configuration portée par la frontière droite. En fait, ces outils de référence donnent l'instruction d'aller chercher un référent qui est, en quelque sorte, enfoui dans le discours précédent. Nous allons proposer un traitement plus fidèle à l'hypothèse de la frontière droite, se basant en priorité sur les contraintes apportées par les informations de surface.

6.3.3 Traitement des structures énumératives à deux temps

6.3.3.1 Structure de surface

Le traitement de la première étape suit les propositions faites dans les sections précédentes et est illustré Figure 6.10. Pour le traitement du second temps, la première possibilité consiste à utiliser le mécanisme de la subordination discursive. La deuxième possibilité consiste à suivre directement les informations fournies par la structure de surface (comme le proposent Egg et Redeker (2008) et Danlos (2008)) et à considérer que la structure sémantique profonde est, en quelque sorte, secondaire et obtenue au niveau de l'interprétation sémantique sans être représentée dans la structure du discours.

Nous retenons cette deuxième solution présentée dans (Vergez-Couret *et al.*, 2008). D'un point de vue cognitif, il nous semble que les informations apportées par la structure énumérative à deux temps est plus saillante que la structure spatio-temporelle de l'histoire, qui vient ultérieurement à l'esprit de l'interpréteur. De ce fait, nous préférons analyser au niveau de l'interprétation sémantique les aspects spatio-temporels de l'histoire plutôt que de les capturer au niveau de la structure du discours.

Cette solution consiste donc à attacher les co-items du deuxième temps directement au constituant correspondant à l'amorce, tel qu'illustré dans la figure 6.15.

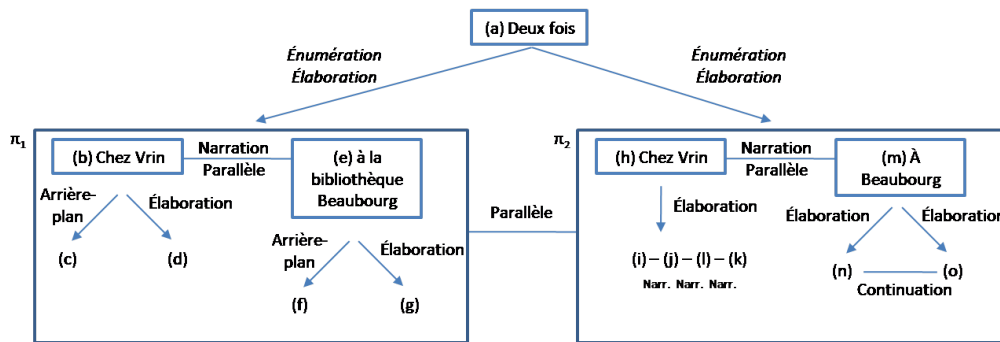


FIG. 6.15 – Structure discursive de (38)

L'amorce devient ainsi le topique discursif principal pour les deux temps de la structure énumérative ainsi que pour les segments (b) & (e) et (k) & (l). La solution adoptée ici est confirmée par l'impossibilité de résoudre des références anaphoriques singulières entre les items et la possibilité d'utiliser un pronom anaphorique pluriel comme ci-dessous :

(42) Chez Vrin, Nil eut peur que les deux amantes se reconnaissent.

En son état actuel, notre proposition demeure descriptive, proposant une structure discursive pour les structures énumératives à deux temps. Elle ne fournit pas de règles permettant de prédire ou contraindre la construction de cette structure. D'abord, la présence d'une structure énumérative classique et l'identification d'un second temps n'est pas suffisante pour prédire que ce second temps présentera la même structure que le premier temps, par exemple un seul des items peut être candidat pour une continuation. Ensuite, nous n'avons pas suffisamment étudié les marqueurs de cette seconde étape. Certains seulement

semblent contraindre suffisamment la structure pour fournir des règles prédictives. Suivant Schnedecker (2001b), il est impossible d'avoir *l'un sans l'autre* tandis que *le premier* peut apparaître sans *le second*. Il semble que la nature prédictive de ce type de structure provienne de la sémantique individuelle des marqueurs d'item plutôt que d'un principe d'organisation général. Il conviendrait donc d'analyser plus en détail ces corrélats anaphoriques. La structure est, par ailleurs, déjà obtenue par la contribution sémantique des marqueurs d'item et non par des règles provenant de structures de plus haut niveau. Les marqueurs d'item dans nos exemples (38) et (39) présentent des variétés intéressantes concernant la façon dont ils sélectionnent des informations nous permettant de comprendre la structure du discours. C'est aussi pour cela que nous ne pouvons pas traiter tous les exemples de structures croisées tels que celui présenté en (40).

6.3.3.2 Structure sémantique profonde

À présent, nous souhaitons combiner la contribution des éléments cohésifs à la surface des textes aux informations sémantiques profondes pour rendre compte de la cohérence globale des structures croisées. Pour ce faire, nous devons analyser la structure sémantique et les implications spatio-temporelles (par exemple la relation temporelle entre les items des deux temps en (38) et la structure alternative/conséquence en (39)). Dans les deux cas, les items du deuxième temps doivent être reliés aux items du premier temps. Cette relation est inférée sur la base d'une identification non ambiguë de l'item concerné par le développement (la reprise du nom propre «Chez Vrin» et «À Beaubourg» en (38) et les corrélats anaphoriques *le premier* et *le second* en (39)) et de l'inférence d'une relation de *Parallèle* entre les constituants complexes correspondant aux deux temps de la structure énumérative.

Notre proposition repose sur le contenu du topique qui domine l'ensemble de la structure énumérative à deux temps. Nous avons vu que ce topique doit inclure les co-items de l'énumération. De plus, les co-items doivent être organisés selon le potentiel des adverbiaux du discours qui les introduisent. Les structures énumératives à deux temps permettent des références anaphoriques au moyen d'outils complexes tel que les corrélats anaphoriques en (39). Ce topique structuré contient également les groupements sémantiques obtenus, par exemple sous les topiques formés par les introducteurs de cadre en (38) ou les marqueurs d'item en (39). Cela suppose que toutes les informations concernant un topique peuvent être accessibles facilement.

Notre proposition présume que le contenu sémantique ne reflète pas directement la cohérence du texte mais que celle-ci est obtenue par des marqueurs de structuration plus explicites. Lorsque le texte ne présente pas de marqueurs de structuration explicites, les informations obtenues au niveau de la sémantique profonde soutiennent la cohérence discursive et correspondent plus directement à la structure du discours.

Nous revenons à présent sur l'identification des introducteurs de cadre du second temps de la structure énumérative.

L'entité plurielle de l'amorce est développée par ces différents éléments constitutifs, correspondant aux co-items. Cependant, les items étant tous les deux disponibles, au moins

avec des reprises anaphoriques au moyen de définis, cela entraîne une ambiguïté. La contribution sémantique des marqueurs d’item du second temps est donc cruciale. Ceux-ci doivent permettre une identification non ambiguë de l’élément qui est développé dans le second temps. Cette ambiguïté est résolue en (38) par la répétition des introducteurs de cadre et en (39) par l’emploi des corrélats anaphoriques *le premier* et *le second*.

En ce qui concerne les introducteurs de cadre, suivant Vieu *et al.* (2005), l’introducteur de cadre introduit dans la structure du discours un constituant semi-explicite pour les introducteurs de cadre des items du premier temps et ceux du second temps. Au constituant semi-explicite introduit par l’introducteur de cadre du premier item du second temps correspond le constituant semi-explicite introduit par l’introducteur de cadre du premier item du premier temps mais celui-ci ne se trouve pas sur la frontière droite (*Cf.* Figure 6.10). Le seul point d’attachement satisfaisant pour les items du second temps est donc le constituant introduit pour la phrase amorce.

Schneidecker (2007) distingue *l’un* et *l’autre* qui sont potentiellement ambigus de *le premier* et *le second* qui résolvent l’ambiguïté. Afin de capturer cette finesse, le topique doit inclure des informations sur l’ordre dans lequel les items sont introduits dans le discours. Les corrélats anaphoriques tels que *le premier*, *dans le premier cas* nécessitent que les items soient ordonnés afin que ce corrélat anaphorique aille chercher le premier. Le point d’attachement, ou plus exactement son contenu sémantique doit inclure une liste ordonnée des items (selon l’ordre dans lequel ils ont été introduit dans le texte) d’où l’importance de notre relation structurelle d’*Énumération* qui rend compte, non pas des aspects sémantiques, mais des aspects textuels. La résolution anaphorique repose ainsi sur la complexité des informations contenues dans ce topique et la sémantique de chaque corrélat anaphorique.

6.3.3.3 Illustration

Nous allons illustrer notre proposition en détaillant les informations spatio-temporelles de (38).

Pour le traitement des introducteurs de cadre, nous suivons les propositions faites dans (Vieu *et al.*, 2005) en postulant qu’un constituant semi-explicite est introduit pour chaque cadre de discours dominant un item. Ce constituant est créé et mis à jour avec les informations sémantiques contenues dans les constituants sous leur portée. En (38), nous avons des cadres de discours spatiaux. Les constituants introduits auront le contenu sémantique suivant :

$$\exists e \exists P(e) \exists \emptyset \wedge P(\emptyset)^5$$

Les constituants introduits par les cadres de discours «Chez vrin» du premier item du premier temps et du premier item du second temps auront donc la forme suivante :

$$\exists e_b \exists P(e_b) \wedge P(\text{Chez Vrin})$$

$$\exists e_h \exists P(e_h) \wedge P(\text{Chez Vrin})$$

⁵*e* pour introduire une éventualité, *P* pour introduire un prédicat correspondant à cette éventualité et \emptyset contient les informations sur la nature du cadre.

P peut être paraphrasé en «ce qui s’est passé», localisé par «Chez Vrin».

Ces informations doivent être exploitées afin d’inférer précisément la structure temporelle de ce discours, donnée ci-dessous :

- (a) L’adverbial temporel «deux fois» introduit un événement complexe e_a composé de deux sous-événements sous-spécifiés e_1 et e_2 avec $e_1 \subset_t e_a$ et $e_2 \subset_t e_a$. Cet événement complexe annonce les deux temps et se trouve partiellement spécifié par «crut sombrer dans le gouffre». Le premier temps correspond à pourquoi il crut sombrer dans le gouffre et le second temps à pourquoi il a *seulement* cru sombrer dans le gouffre. Mais cette annonce des deux temps repose sur des indices très fins qui sont difficilement exploitables.
- (b) Comme nous l’avons expliqué ci-dessus, les cadres de discours introduisent des événements sous-spécifiés. Pour éviter la profusion des niveaux de représentation non nécessaires, comme expliqué dans (Asher *et al.*, 2008), nous cherchons s’il existe déjà des éventualités introduites et vérifions si elles correspondent aux éventualités nouvellement introduites. Dans notre cas, nous avons déjà e_1 et e_2 qui sont des événements sous-spécifiés en attente de spécifications. Nous avons donc $e_1 = e_b$. Nous gardons à l’esprit que l’histoire ayant eu lieu chez Vrin se déroule en deux temps (b) et (h). La relation entre ces éventualités est la suivante : $e_b \sqcap e_h = e_1$. Les deux éventualités e_b et e_h réunies forment l’éventualité e_1 .
- (c-d) Les informations aspectuelles, i.e. l’attribution entre l’état en (c) et l’événement en (d) ainsi que le marqueur *quand* permet d’inférer la relation d’*Arrière-plan* et la relation temporelle correspondante $s_c \circ e_d$ et par raisonnement temporel $s_c \subset_t e_b$ et $e_d \subset_t e_b$.
- (e) De la même façon que pour le segment (b), nous obtenons $e_e = e_2$. Le marqueur *ensuite* permet de plus l’inférence de la relation de *Narration* entre les deux constituants π_b et π_e , et de la relation temporelle correspondante $e_b <_t e_e$. Enfin, nous gardons à l’esprit que l’histoire se déroulant à Beaubourg a lieu en deux temps (e) et (m). La relation entre ces éventualités est la suivante : $e_e \sqcap e_m = e_2$. Les deux éventualités e_e et e_m réunies forment l’éventualité e_2 .
- (f-g) De la même façon que pour (c-d), nous obtenons $s_f \circ e_g$ et par raisonnement temporel $s_f \subset_t e_e$ et $e_g \subset_t e_e$.
- (h) Le nouveau cadre de discours correspond à celui donné en (b). e_b et e_h renvoient au même cadre spatial. Cela nous aide à considérer les relations temporelles s’établissant entre les sous-éventualités décrites sous (b) et sous (h).
- (i-j) Dans un premier temps, le marqueur *dès que* déclenche l’inférence de *Narration* entre (i) et (j) et l’inférence de la relation temporelle correspondante $e_i <_t e_j$. Ensuite, deux indices indiquent que l’événement e_i entretient des relations potentielles avec le discours précédent : la co-référence entre e_h et e_b ainsi que l’emploi du plus-que-parfait (au lieu du passé simple). Le plus-que-parfait signale une relation temporelle spécifique positionnant l’événement e_i comme antérieur à un autre événement disponible. De plus, la combinaison d’informations liées à la sémantique lexicale et à des connaissances du monde permettent d’inférer une relation similaire à celle exprimée par le prédicat *Occasion* entre les événements «Laure entra» et «Nil avait vu Laure». Nous proposons ici de tenir compte de ces relations entre éventualités au

niveau de l'interprétation spatio-temporelle profonde plutôt qu'au niveau rhétorique pour lequel les informations saillantes sont données par les introducteurs de cadre. Autrement dit, nous avons suffisamment d'informations pour inférer l'ordre temporel des événements en (d) et (i) mais la question de leur attachement n'est pas résolue au moyen de règles de construction du discours. De ce fait, nous n'inférons pas ici la relation de *Narration*.

- (j-k-l) Un raisonnement similaire fondé sur les informations liées à la sémantique lexicale et aux connaissances du monde, similaire à celui utilisé pour (i) et (j), permet d'identifier une relation de type *Occasion* entre «se cacher» dans (j) et «sortir» dans (k). Mais dans ce cas, l'attachement de (j) à (k) est considéré et *Narration* est inférée. Enfin, la construction avec *après que* permet l'inférence de la relation de *Flashback*. Les relations temporelles correspondantes aux relations de *Narration* et de *Flashback* permettent d'inférer $e_j <_t e_l <_t e_k$.
- (m) Comme pour (b) et (h), le nouveau cadre temporel introduit en (m) correspond à celui introduit en (e), nous aidant à considérer les relations temporelles s'établissant entre les sous-éventualités décrites sous (e) et sous (m).
- (n-o) Les éventualités décrites en (n) et (o) sont des états. Ils se superposent et sont tous les deux inclus dans e_m . Cette inclusion nous amène à considérer les relations temporelles potentielles avec les autres éventualités de l'histoire se déroulant à Beaubourg, décrite en (e). Nous pouvons à nouveau inférer un lien similaire à *Occasion* entre «Anne-geneviève apparut» et «Anne-Geneviève n'est restée qu'un instant». Nous inférons donc la relation temporelle correspondante $e_g < e_n$.

Les relations temporelles décrites ci-dessus sont schématisées Figure 6.16.

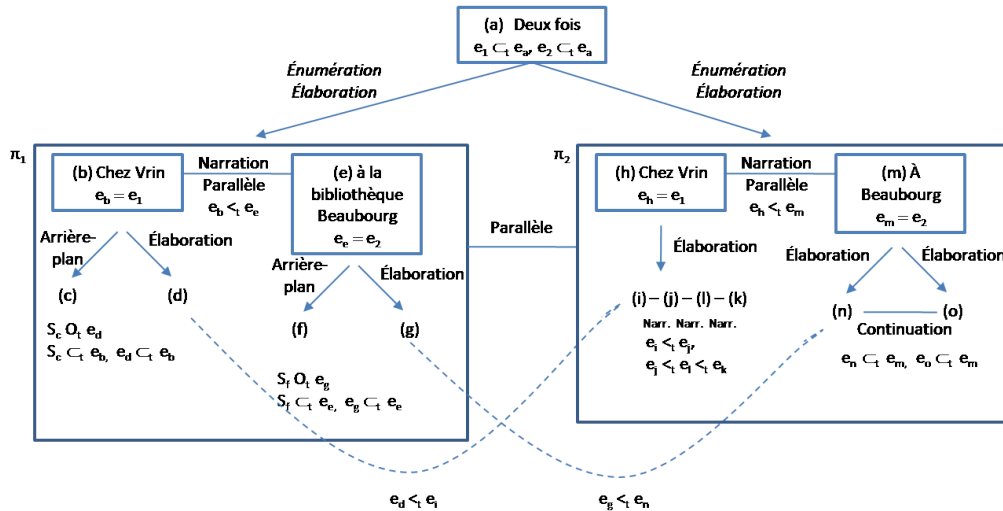


FIG. 6.16 – Structure temporelle de (38)

6.3.3.4 Généralisation

Bien que satisfaisant pour un exemple tel que (38), le traitement proposé ci-dessus est bien trop spécifique pour être généralisable.

Nous dégageons, à présent, les similitudes présentes en (38) et (39) afin de proposer des généralisations pour leur traitement. Les deux exemples présentent :

- (i) une entité de discours plurielle : «deux fois» en (38) et «deux options» en (39). Dans les deux cas, cette entité est développée formant une structure *via* la relation d'Énumération. Les items du premier temps sont reliés par une relation coordonnante, *Narration* en (1) et *Alternation* en (2).
- (ii) des marqueurs du discours explicitant l'organisation discursive : les introducteurs de cadre en (38) et les corrélats anaphoriques en (39) et un parallélisme structurel entre les constituants complexes qui signalent une structure de plus haut niveau. L'idée sous-tendant la subordination discursive (Asher, 1993) est reprise ici mais d'une façon plus conservatrice vis-à-vis du principe de la frontière droite.

Nous proposons de regarder ce qu'il se passe lorsque des constituants complexes sont reliés par des relations de contenu, *Narration* en (38) et *Conséquence* en (39), Cf. respectivement Figures 6.17 et 6.18.

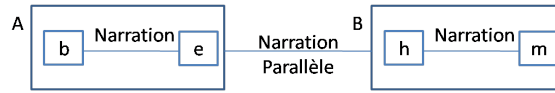


FIG. 6.17 – Constituants complexes en (38)

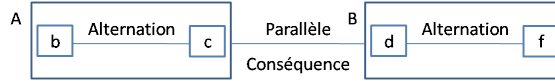


FIG. 6.18 – Constituants complexes en (39)

En SDRT, les constituants complexes sont des emballages organisateurs (en formant des groupements de segments minimaux) qui ne jouent pas un rôle direct sur les contraintes sémantiques. Ils sont des produits de la structure créée par l'interprétation sémantique. Dans le cas des structures énumératives à deux temps, nous voudrions autoriser l'attachement de ces constituants complexes par des relations de contenu et nous proposons que les contraintes sémantiques de ces relations soient vérifiées entre les constituants formant le constituant complexe de la façon suivante :

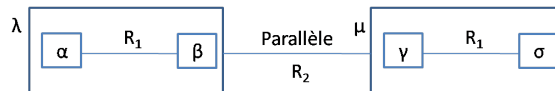


FIG. 6.19 – Généralisation

$$R_1(\alpha, \beta, \lambda) \wedge R_1(\gamma, \delta, \mu) \wedge \text{Parallèle}(\lambda, \mu, \nu) \wedge \emptyset_{R_2}(\alpha, \gamma) \wedge \emptyset_{R_2}(\beta, \delta) \rightarrow R_2(\lambda, \mu, \nu)$$

L'application de ces contraintes aux exemples (38) et (39) donne les résultats suivants :

$$\text{Narration}(b, e, A) \wedge \text{Narration}(h, m, B) \wedge \text{Parallèle}(A, B, 0) \wedge \emptyset_{\text{Narration}}(b, h) \wedge \emptyset_{\text{Narration}}(e, m) \rightarrow \text{Narration}(A, B, 0)$$

$$\text{Alternation}(b, c, A) \wedge \text{Alternation}(d, f, B) \wedge \text{Parallèle}(A, B, 0) \wedge \emptyset_{\text{Conséquence}}(b, d) \wedge \emptyset_{\text{Conséquence}}(c, f) \rightarrow \text{Conséquence}(A, B, 0)$$

Pour l'attachement de cette structure (Cf. Figure 6.18) à la phrase amorce, en (39), nous proposons qu'une double relation d'*Énumération* et d'*Élaboration* soit inférée entre l'amorce et le premier temps de la structure énumérative, le second temps étant seulement attaché par *Énumération*. Il serait tentant d'ajouter la relation, s'établissant entre l'amorce et le premier temps, entre l'amorce et le second temps. En effet, le premier temps étant une élaboration de l'amorce, la fonction sémantique de ces éléments doit être équivalente à celle du second temps. Mais le second temps n'élabore plus les options, il donne les conséquences de celles-ci.

6.3.4 Synthèse

Pour résumer les avantages de notre traitement :

- i. c'est un traitement standard plus conservateur que la subordination discursive (Asher, 1993) vis-à-vis de la frontière droite sur un sujet qui n'a pas été beaucoup discuté depuis cette proposition ;
- ii. ce traitement s'appuie sur les observations empiriques sur les corrélats anaphoriques de Schnedecker (2007) ;
- iii. et enfin il ouvre un cadre qui peut être étendu à des cas plus généraux.

Cette étude des structures énumératives a mis en avant le niveau textuel, qui prime dans l'interprétation des structures énumératives classiques et des structures énumératives à deux temps.

Ces structures sont très riches pour repérer la relation d'*Élaboration* ou de *Topique*. Elles ont permis de mettre au jour un certain nombre d'indices tels que les cadres de discours et les corrélats anaphoriques.

6.4 Conclusion

Nous avons montré les fonctionnements discursifs différents des marqueurs structurels *d'abord*, *dans un premier temps* et *premièrement*. *D'abord* apparaît fréquemment dans des contextes d'*Élaboration* mais son fonctionnement général est de signaler une relation subordonnante. *Dans un premier temps* signale toujours une *Élaboration* mais la relation est inférée soit avec un topique explicite, soit avec un topique implicite. Nous verrons dans le chapitre 9 les difficultés liées à la gestion des topiques implicites dans le cadre d'une annotation manuelle des relations de discours. Enfin, *premièrement* présente plus de similitudes avec *d'abord* qu'avec *dans un premier temps*. Ces trois adverbes apparaissent fréquemment

dans des structures énumératives. Une analyse de ces structures avec des relations de discours à l'œuvre fait intervenir des faisceaux d'indices basés sur la nature de l'amorce, la nature des marqueurs d'item et leurs interactions.

Notre proposition repose sur l'introduction d'une nouvelle relation de discours *Énumération* qui rend compte du niveau textuel et non du niveau sémantique. Nous justifions l'introduction de cette nouvelle relation de discours par le fait que la reconnaissance de la structure sur la base d'éléments à la surface du texte (amorce et marqueurs d'item) est nécessaire pour rendre compte efficacement de l'interprétation de la structure sémantique. Nous considérons ainsi les structures énumératives comme des structures qui déclenchent l'inférence d'une relation de contenu. Les relations de discours associées aux structures énumératives peuvent être des élaborations, des explications, des conséquences, des résultats... Le repérage automatique de ces structures énumératives peut nous permettre d'accéder à des «lieux propices» aux élaborations.

Chapitre 7

Une structure syntaxique, le gérondif

- Vous dessinez ou vous m'écoutez ?
- Je vous écoute en dessinant.
(Ensemble c'est tout, Anna Gavalda, 2004)

Sommaire

7.1	La construction gérondive	208
7.1.1	Aspects syntaxiques	208
7.1.1.1	Forme verbale à fonction adverbiale	208
7.1.1.2	Points d'attachement du syntagme gérondif	209
7.1.1.3	Place du syntagme gérondif	210
7.1.2	Aspects sémantiques	210
7.2	La typologie de Halmøy (1982)	211
7.2.1	Repère temporel (RT)	211
7.2.2	Relation de cause, de condition ou de moyen (A)	211
7.2.2.1	<effet, cause>	212
7.2.2.2	<conséquence, condition>	212
7.2.2.3	<résultat, moyen>	212
7.2.3	Relation d'inclusion ou d'équivalence (A')	213
7.2.4	Relation de concomitance (B)	214
7.2.5	Relation d'hyponymie (B')	216
7.3	Représentation discursive du gérondif	218
7.3.1	Configuration A <résultat, moyen>	218
7.3.1.1	Moyen ₁	221
7.3.1.2	Moyen ₂	222
7.3.1.3	Bilan	222
7.3.2	Configuration A'	223
7.3.3	Configuration B	224
7.3.4	Configuration B'	228
7.4	Conclusion	229

Les travaux de Fabricius-Hansen et Behrens (2001) nous ont mise sur la voie de cette étude du gérondif. Ces dernières ont proposé une approche comparative (entre langues -

allemand, norvégien, anglais) des marqueurs et structures utilisés pour exprimer la relation d'*Élaboration*. Leur point de départ est le marqueur allemand *indem*, conjonction de subordination introduisant une subordonnée de manière ou de moyen ou une subordonnée temporelle. Elles notent que, généralement, ce marqueur signale une *Élaboration* entre la principale et la proposition subordonnée. Selon les auteurs, l'étude de ce marqueur et de sa traduction en norvégien et en anglais permet d'approfondir les connaissances sur la relation d'*Élaboration*. En français, la proposition subordonnée introduite par *indem* est généralement traduite au moyen d'un gérondif. Dans le corpus OMC¹, 34 *indem* sur 43 du corpus original allemand sont traduits en français par un gérondif. Mais comme nous le verrons dans ce chapitre, toutes les propositions gérondives (tout comme les propositions subordonnées introduites par *indem*) ne sont pas des élaborations de la proposition régissante (ou rectrice). Nous allons étudier à partir des travaux de Halmøy (1982, 2003), et plus particulièrement de la typologie des constructions gérondives qu'elle propose, les relations discursives et temporelles s'établissant dans ce type de construction. De ce fait, nous allons nous intéresser exclusivement à des relations s'établissant entre deux propositions à l'intérieur d'une même phrase.

Le gérondif n'est pas la seule construction syntaxique permettant d'exprimer des élaborations. Le participe présent peut également avoir cette fonction. Dans le cadre de ce chapitre, nous devons renoncer à une analyse commune, par manque de temps et de place, des deux constructions, qui par ailleurs ne sont pas toujours interchangeables et présentent des fonctionnements discursifs différents.

Cette étude des constructions gérondives invite à une réflexion sur le statut des éventualités dans ce type de construction, l'éventualité dénotée par la principale et celle dénotée par le gérondif et ses compléments : dans quels cas les éventualités réfèrent-elles à la même éventualité, à des éventualités différentes, à des éventualités concomitantes ? dans quels cas le syntagme gérondif est-il une élaboration de la proposition régissante ?

Dans ce chapitre, nous commençons par présenter les aspects syntaxiques et sémantiques généraux du gérondif (section 7.1) et la typologie proposée par Halmøy (2003) afin de montrer l'éventail des interprétations possibles avec cette construction (section 7.2). Puis nous basant sur cette typologie, nous allons déterminer les configurations qui nous paraissent relever de la relation d'*Élaboration* et proposer une analyse discursive de ces configurations (section 7.3).

7.1 La construction gérondive

7.1.1 Aspects syntaxiques

7.1.1.1 Forme verbale à fonction adverbiale

Le **gérondif** est une forme non conjuguée du verbe. C'est une forme verbale, admettant toutes les expansions du verbe (compléments et modifieurs) dont le sujet n'est pas exprimé. La règle normative établie par les grammaires traditionnelles spécifie que ce sujet doit être co-référent au sujet de la proposition régissante. Halmøy (2003) signale que de nombreux exemples, empruntés autant au français contemporain qu'aux auteurs classiques, défient

¹Olso Multilingual Corpus.

cette règle, comme en atteste le proverbe suivant :

- (1) L'appétit vient en mangeant.

D'autre part, le **syntagme gérondif** (le gérondif et ses compléments ou modificateurs) présente le fonctionnement d'un complément adverbial (ou d'un complément circonstanciel dans la terminologie de la grammaire traditionnelle). Il en a le comportement et la distribution. Celui-ci peut ainsi être coordonné avec un syntagme prépositionnel en fonction adverbiale :

- (2) Je voyage *dans les livres* et ***en allant*** *au cinéma*. (Halmøy, 2003, p. 80)

En tant que forme adverbiale, le syntagme gérondif est une forme régie, une expansion ou un modifieur, généralement du prédicat verbal ou d'une phrase. Il est facultatif : il peut donc être supprimé sans affecter la grammaticalité de la phrase.

- (3) Il est parti en courant
Il est parti. (Halmøy, 2003, p. 70)

La **construction gérondive** est le couple formé de la proposition principale et du syntagme gérondif.

7.1.1.2 Points d'attachement du syntagme gérondif

Halmøy (2003) propose un inventaire des points d'attachement du syntagme gérondif. Nous le résumons ci-dessous.

Dans les cas typiques, le syntagme gérondif est rattaché à un prédicat verbal. Le verbe régissant le syntagme gérondif peut être fini (à tous les temps et tous les modes de toutes les personnes) ou non-fini, par exemple :

- (4) Elle est sortie de la ville en longeant la mer (Halmøy, 2003, p. 72)
(5) Sortir en longeant la mer (Halmøy, 2003, p. 72)

Dans certains cas, plus rares, le syntagme gérondif est attaché :

- à un adverbe ou une locution adverbiale

- (6) Le lendemain matin, en buvant son chocolat, elle relit ses notes, consulte ses documents, et entreprend la rédaction du texte. (Halmøy, 2003, p. 73)

Le syntagme gérondif «en buvant son chocolat» précise un moment dans l'ensemble défini par la locution adverbiale «le lendemain matin». L'ensemble «le lendemain matin, en buvant son chocolat» constitue un repère temporel pour l'ensemble de la proposition principale.

- à un syntagme nominal

- (7) L'anglais en s'amusant (Halmøy, 2003, p. 75)

Le gérondif dépend certes syntaxiquement du SN «l'anglais» mais le prédicat verbal «apprendre l'anglais» nous semble disponible pour l'interprétation.

Enfin, dans certains cas, le syntagme gérondif n'a pas de point de rattachement explicite :

- (8) En attendant Godot (Beckett) (Halmøy, 2003, p. 76)

Il est ici sous-entendu «tout ce qui se passe en attendant Godot».

Dans de nombreux cas, les gérondifs ne se laissant pas rattacher à un constituant de la phrase sont des syntagmes gérondifs grammaticalisés. On peut citer à titre d'exemple des figements adverbiaux *en attendant*, *en passant*, des figement prépositionnels *en partant de*, *en passant par* et des figements conjonctionnels *en admettant que*, *en supposant que*.

7.1.1.3 Place du syntagme gérondif

Le syntagme gérondif est à première vue mobile et peut se placer à toutes les césures de la phrase. Cette remarque est valable pour les syntagmes gérondifs à valeur de repère temporel (Cf. section 7.2 pour la présentation de la typologie de Halmøy (2003)). Concernant les autres types de gérondifs, la position du syntagme est contrainte par le sémantisme de la construction et la répartition thématique des constituants, comme nous le verrons au fur et à mesure de la présentation des différentes configurations des constructions gérondives décrites par Halmøy.

7.1.2 Aspects sémantiques

Les grammaires traditionnelles considèrent la valeur adverbiale (ou circonstancielle) du gérondif. Dans la *Grammaire Méthodique du Français*, on relève la considération suivante :

Il équivaut à un adverbe et assume la fonction d'un complément circonstanciel de manière, de moyen, de temps, de cause, de condition ou d'opposition, selon la relation avec le reste de la phrase. (Riegel et al., 1994, p. 342)

Halmøy (2003) souligne l'inadaptation de ces catégories de circonstanciels pour traiter les différentes configurations du syntagme gérondif. Elle propose donc une typologie des constructions gérondives en étudiant les relations s'établissant entre le verbe régissant² et le syntagme gérondif. Ces relations naissent du rapprochement du sémantisme de deux verbes et de leurs expansions, le gérondif étant une forme non marquée. La citation ci-dessus résume très clairement la description du gérondif proposée par Halmøy (1982) :

Le gérondif établit un rapport de subordination syntaxique entre deux verbes : deux procès sont ainsi mis en rapport, le VP étant syntaxiquement subordonnant, le gérondif étant subordonné. Selon les verbes en jeu, cette subordination syntaxique peut prendre diverses valeurs sémantiques, mais le gérondif en tant que tel est non marqué quant au rapport sémantique de la relation qu'il établit. La forme en r(V)ant à elle seule ne donne aucune indication sur la valeur sémantique de cette relation, mais elle permet un certain éventail d'interprétations, qui dépendent de la combinaison des deux racines verbales ainsi rapprochées. L'interprétation du sens de la relation ne peut se faire qu'a posteriori, et est déterminée par le jeu des deux racines verbales en question. D'autres

²Ou verbe recteur.

éléments du contexte linguistique ou situationnel ou pragmatique concourent à préciser la nature sémantique de ce rapport. (Halmøy, 1982, p. 252)

Dans la section suivante, nous allons présenter la typologie de Halmøy (2003). Nous nous baserons ensuite sur sa terminologie pour notre propre analyse en relations de discours.

7.2 La typologie de Halmøy (1982)

La typologie de Halmøy (2003) se divise en cinq configurations regroupées en trois catégories, rendant compte de la variété interprétative de la construction gérondive :

- les gérondifs «repère temporel» (RT)
- les gérondifs exprimant une relation logiquement orientée (A et A')
- les gérondifs exprimant une concomitance entre deux procès (B et B')

Plus spécifiquement, les gérondifs des catégories A' et B' précisent un aspect du procès exprimé par le verbe régissant. Nous illustrons chacune de ces configurations dans les sous-sections suivantes.

7.2.1 Repère temporel (RT)

Dans cette première catégorie, sont classées les constructions gérondives où seule une interprétation temporelle est possible.

- (9) **En approchant** de la chambre, j'entendis les cris douloureux d'un homme malgré la porte fermée. (Halmøy, 2003, p. 92)
- (10) **En rentrant** dans la chambre, il se rendit compte qu'elle n'avait été tirée d'embarras que pour un embarras plus grand encore. (Halmøy, 2003, p. 92)
- (11) Une femme a toujours un paquet de linge à déposer **en partant** au cinéma, le pain à ne pas oublier **en rentrant** du travail. (Halmøy, 2003, p. 93)

Le gérondif «repère temporel» est très fréquent, notamment dans les oeuvres de fiction. Cette fonction du gérondif a été étudiée par Gettrup (1977) qui souligne que le rôle du syntagme gérondif est «*d'indiquer un moment ou un espace de temps qui sert de repère temporel à l'action du verbe fini*».

D'après Halmøy (2003), le gérondif est généralement un verbe de déplacement («en approchant», «en partant», «en rentrant») et le verbe régissant un verbe de perception («entendis») ou un prédicat intellectuel («se rendit compte»). Dans cette configuration, la position du gérondif est très libre, celui-ci pouvant se situer à chaque césure de la phrase.

7.2.2 Relation de cause, de condition ou de moyen (A)

Dans la configuration dite «configuration A», il existe une relation de dépendance entre le gérondif et le verbe régissant. Cette dépendance <proposition régissante, syntagme gérondif> peut prendre trois formes, que l'on peut résumer sous la forme des couples :

- <effet, cause>,
- <conséquence, condition> ou
- <résultat, moyen>.

L'ordre dans lequel ces couples se manifestent est libre. Selon Halmøy, le point commun de ces trois configurations est d'ordre temporel : le point de départ du procès dénoté par le syntagme gérondif (SG) est antérieur au procès dénoté par le verbe régissant (VR). La relation est ainsi toujours orientée, schématisée de la façon suivante par Halmøy (1982) : $SG \rightarrow VR$. Halmøy parle d'*antériorité logique* du procès décrit par le syntagme gérondif, qui précède toujours celui décrit par le verbe régissant.

7.2.2.1 <effet, cause>

Une interprétation causale s'ajoute à l'interprétation de repère temporel.

- (12) Marion a pleuré **en m'entendant** crier. (Halmøy, 2003, p. 95)
- (13) Alors, **en ne signant pas**, vous retardez la sépulture de votre fils d'autant. (Halmøy, 2003, p. 95)

En (12), c'est lorsque/parce que/après que Marion l'a entendu crier qu'elle s'est mise à pleurer.

Le gérondif est neutre quant à l'interprétation plutôt temporelle ou plutôt causale du couple <verbe régissant, syntagme gérondif> : l'interprétation causale naît de la relation existant entre le sémantisme des structures prédicats/arguments en jeu et des connaissances du monde.

7.2.2.2 <conséquence, condition>

- (14) **En partant** de bonne heure, nous pouvons être rentrés dans la nuit. (Halmøy, 2003, p. 96)
- (15) Il se porterait mieux **en mangeant** moins. (Halmøy, 2003, p. 96)

L'interprétation conditionnelle peut être déclenchée par divers facteurs, tels que le sémantisme du verbe régissant dénotant une possibilité («pouvons») ou le conditionnel («se porterait»).

7.2.2.3 <résultat, moyen>

- (16) Il essayait de se réchauffer les pieds **en les frottant** contre ses mains. (Halmøy, 2003, p. 97)
- (17) **En prêtant** l'oreille, je parvenais à discerner quelques bribes de leur conversation. (Halmøy, 2003, p. 97)

Les gérondifs exprimant le moyen par lequel le procès décrit par le verbe régissant est obtenu sont souvent considérés comme les cas prototypiques de la construction. Le moyen est décrit dans le Nouveau Petit Robert (Édition 2008) comme «ce qui sert pour arriver à une fin».

L'interprétation du moyen est favorisée par la présence d'un verbe régissant exprimant une idée d'effort ou d'intention («essayait de») ou de résultat («parvenais à»).

Les gérondifs répondent à la question «comment?» :

- (18) Comment essayait-il de se réchauffer les pieds ? En les frottant contre ses mains.
 (19) Comment parvenais-je à discerner quelques bribes de leur conversation ? En prêtant l'oreille.

Halmøy (1982) signale qu'il existe une relation d'antériorité temporelle du procès décrit par le syntagme gérondif qui est mise en avant par la manipulation suivante :

- (20) Il prêtait l'oreille et parvenait à discerner quelques bribes de leur conversation.
 * Il parvenait à discerner quelques bribes de leur conversation et prêtait l'oreille.
 (21) Il se frottait les pieds contre ses mains et essayait de se les réchauffer.
 * Il essayait de se réchauffer les pieds et se les frottait contre ses mains.

La coordination avec *et* est possible à condition que l'ordre temporel des deux procès soit respecté.

Dans les derniers cas de gérondif-moyen qu'Halmøy (2003) considère, l'interprétation de moyen se superpose à une autre interprétation qu'elle appelle *relation d'inclusion*.

- (22) Nous essayons de satisfaire les besoins des étudiants **en leur mettant** dans la main des documents qu'ils utiliseront dans leur pratique professionnelle. (Halmøy, 2003, p. 98)

L'exemple (22) nous paraît tout à fait similaire à l'exemple (16). Halmøy (2003, p. 98) note à propos de cet exemple : «*En effet, on peut considérer que le procès exprimé par le gérondif est en quelque sorte «inclus» dans le procès exprimé par le VR³ : «mettre les documents en question dans les mains des étudiants, c'est une façon, un moyen parmi d'autres⁴, d'essayer de satisfaire leurs besoins»...*».

Cette relation d'inclusion, employée par Halmøy, n'est pas une inclusion temporelle mais plutôt une sorte d'inclusion de sens : le procès exprimé par le verbe régissant peut faire intervenir un ensemble de moyens et le syntagme gérondif dénote un de ces moyens.

La configuration suivante A' rend compte des gérondifs dont l'interprétation fait intervenir une relation d'inclusion avec une atténuation quelquefois complète de l'interprétation de moyen.

7.2.3 Relation d'inclusion ou d'équivalence (A')

- (23) Ségolène Royal avait volé dans les plumes de ses collègues **en annonçant** une réforme de l'accouchement sous X. (Halmøy, 2003, p. 99)
 (24) Aucun hôpital n'accepterait de violer la loi, **en mettant au monde** l'enfant d'un couple non marié. (Halmøy, 2003, p. 99)

Par rapport aux constructions gérondives <résultat, moyen>, l'interprétation de moyen s'atténue. La proposition régissante a la caractéristique d'être abstraite et/ou métaphorique

³Verbe Régissant.

⁴Notre mise en valeur.

et le rôle du syntagme gérondif est de donner une réalisation concrète venant spécifier le contenu de la proposition régissante. Halmøy (2003) considère le verbe régissant comme hyperonymique dans la mesure où le syntagme gérondif décrit une réalisation concrète du procès dénoté par le verbe régissant. En nous basant sur des connaissances du monde, nous savons que «annoncer une réforme de l'accouchement sous X» est une façon de «voler dans les plumes de ses collègues»; «mettre au monde l'enfant d'un couple non marié» est une façon de «violer la loi». Les prédicats («voler dans les plumes», «violer la loi») peuvent être accomplis de plusieurs façons et le syntagme gérondif (respectivement «annoncer une réforme de l'accouchement sous X» et «mettre au monde l'enfant d'un couple non marié») en dénotent une. Le procès décrit par le syntagme gérondif est de ce fait une manifestation particulière du procès dénoté par le verbe régissant. Cependant, nous notons qu'avec la construction gérondive, il semble que l'interprétation d'équivalence temporelle prime toujours.

Comme le signale Halmøy (2003), il existe un continuum entre les interprétations de moyen (de la configuration A) et les interprétations d'inclusion (de la configuration A'). Une analyse en relations de discours des constructions gérondives de ces deux configurations nous conduira à un autre découpage (section 7.3).

7.2.4 Relation de concomitance (B)

La configuration B se distingue principalement par le fait que la relation s'établissant entre le verbe régissant et le syntagme gérondif n'est plus une relation orientée.

- (25) Le chef faisait les cents pas **en fumant** sa pipe en bambou. (Halmøy, 2003, p. 101)
- (26) Je vous écoute **en dessinant** (*Ensemble c'est tout*, Anna Gavalda)

Les deux procès décrits se déroulent dans le même temps sans qu'il existe un lien logique entre eux. La relation de concomitance ainsi établie est fortuite. Dans cette configuration, le gérondif peut être précédé de *tout* comme en (27).

- (27) À ses côtés, Jacques Chirac envoyait des baisers à la foule **tout en aplatisant** une mèche folle, dérangée par la brise. (Halmøy, 2003, p. 102)

Le verbe régissant et le syntagme gérondif semblent interchangeable : on peut aussi bien «faire les cent pas en fumant sa pipe» que «fumer sa pipe en faisant les cent pas»; mais cette manipulation entraîne, comme le souligne Gettrup (1977), un changement de perspective.

- (28) Paul chante **en se rasant**.
- (28)' Paul se rase **en chantant**.

Dans les deux cas, les procès se déroulent de façon simultanée. Gettrup (1977) décrit la différence de perspective de ces deux exemples en terme de nouveauté : en (28), le gérondif est un fait connu et en (28)', un fait non connu. La notion de connu/non connu est expliquée par Gettrup (1977) par le fait que la plupart des hommes se rasent régulièrement et de ce fait la mention de cette action n'a rien d'insolite et est considérée comme connue. Il

parle d'*acquis existentiel* lorsque cette notion repose sur des faits extra-linguistiques. En (28), le syntagme gérondif fait office de repère temporel tandis qu'en (28)' il fait office de circonstance concomitante ou circonstance accompagnante (*accompanying circumstance* introduit par Fabricius-Hansen et Behrens (2001) pour parler de cas similaires en anglais, norvégien et allemand. Gettrup (1977) note que l'interprétation de repère temporel du gérondif est inversement proportionnelle au degré de nouveauté de celui-ci.

L'interprétation temporelle de (28) peut être mise en évidence par la possibilité de faire une paraphrase du syntagme gérondif au moyen de *quand* tandis que la même paraphrase n'est pas satisfaisante dans les cas de circonstance accompagnante comme en (28)' :

(29) Je chante en me rasant → Quand je me rase, je chante.

(30) Je me rase en chantant ↗ Quand je chante, je me rase.

Le changement de perspective est utilisé à des fins humoristiques dans la blague suivante⁵ :

(31) Un jésuite et un Franciscain se trouvent ensemble dans un train. Tous deux fument la pipe en lisant. Arrive l'heure de l'office. Le Franciscain éteint sa pipe, prend son bréviaire et commence à prier. De son côté, le Jésuite prend aussi son bréviaire et commence également à prier, mais il garde sa pipe à la bouche et continue à tirer dessus. Le franciscain s'en aperçoit et dit un peu scandalisé :

- Mais, mon cher confrère, je vois que vous continuez à fumer. Pourtant, moi, quand j'ai demandé à mon supérieur si l'on pouvait **fumer en priant**, il m'a dit : «Bien sûr que non.»

- Ah, répond le jésuite, mais moi, je lui ai demandé au contraire si l'on pouvait **prier en fumant**, et il m'a dit : «Mais bien sûr, on peut prier en toutes circonstances !» (<http://www.puiseralasource.org/index.php?page=humour>)

Le caractère présupposé des gérondifs RT, comme celui des subordonnées temporelles (Le Draoulec, 1999), est mis en évidence avec les manipulations suivantes (Kleiber, 2008) :

(32) Le franciscain fume en priant.

Le franciscain ne fume pas en priant. → Le franciscain prie.

(33) Le jésuite prie en fumant.

Le jésuite ne prie pas en fumant. ↗ Le jésuite fume.

Lorsque le franciscain demande à son supérieur s'il peut «fumer en priant», il reçoit une réponse négative tandis que le jésuite en demandant s'il peut «prier en fumant» reçoit une réponse positive. Le supérieur du jésuite a justement interprété «en fumant» comme une circonstance accompagnante plutôt que comme un repère temporel de la prière.

L'apparente réversibilité de la construction est liée au fait que les deux procès sont dans une relation non orientée. Les deux procès sont concomitants et la co-référence du sujet est, dans cette configuration, une condition *sine qua non*. Cependant, l'inversion de la construction entraîne un changement de perspective, lié à un changement au niveau de la progression thématique.

⁵Nous reprenons ici une version modifiée de la blague d'abord mentionnée dans la thèse de B. Moortgat (1978).

Gosselin (1996) propose une étude de la temporalité du français en décrivant les relations temporelles en termes d'intervalles et de bornes d'intervalle dans sa théorie des relations référentielles. Pour décrire les cas de la configuration B (ainsi que B' présentée dans la section suivante), il parle de *co-appartenance à une même vue*, marquée par la coïncidence des intervalles de référence, impliquant la simultanéité temporelle des procès, comme dans l'exemple suivant :

(34) Ce fut **en jurant** qu'il mangea son chocolat dans la casserole. (Amourette, 2006)

L'idée que nous retenons prioritairement est la co-appartenance à une même vue ou à une même scène. Il nous semble effectivement que les deux éventualités décrites par le verbe régissant et le syntagme gérondif décrivent en deux propositions une éventualité complexe. Nous y reviendrons dans la section dédiée à l'analyse des constructions gérondives avec les relations de discours.

Auparavant, nous allons présenter la dernière configuration B', faisant intervenir la relation d'hyponymie.

7.2.5 Relation d'hyponymie (B')

Dans ce cas particulier, la construction gérondive n'exprime, d'après Halmøy (2003), qu'une seule action.

(35) Elle répondit **en bafouillant** que c'était la fille d'une ancienne voisine.

(36) Ludi vient à leur rencontre **en galopant**.

Le sujet du verbe régissant et du gérondif doit, de ce fait, être obligatoirement co-référent. Dans cette configuration, le verbe régissant peut être un verbe de dire, en (35), ou un verbe de déplacement ou de mouvement, en (36), et le gérondif, intraprédicatif, est en relation d'hyponymie avec le verbe régissant. Selon Halmøy (2003), la nomination de *manière* n'est adéquate que pour ces gérondifs hyponymes. Il faut donc bien entendu distinguer les deux configurations suivantes :

(37) partir en boitant.

(38) partir en haussant les épaules.

Alors qu'«en boitant» permet de dénoter la façon dont s'effectue le déplacement supposé par l'action de «partir», «en haussant les épaules» n'est pas une façon de «partir» et n'est donc pas considéré comme étant la manière d'accomplir cette action.

Halmøy (2003) observe que le gérondif a une valeur rhématique et que le verbe de déclaration (dire dans les cas prototypiques) est non seulement thématique, mais redondant. Ainsi «dit-il en bafouillant» serait équivalent à «bafouilla-t-il». Ce qui ne nous semble pas, dans tous les cas, tout à fait exact : en (35), le verbe de dire «répondre» apporte des indications sémantiques supplémentaires au seul fait qu'il y ait eu acte de parole. La proposition «elle répondit en bafouillant» implique «elle bafouilla», mais ne lui est pas équivalente.

Si la relation d'hyponymie relie «dire» et «bafouiller», qui est une façon de dire, elle ne relie pas directement «répondre» et «bafouiller». On peut inférer cette relation en passant par le fait que «répondre» constitue un acte de dire. Cette remarque est aussi valable pour les verbes de déplacement. «En galopant» dénote une manière de se déplacer et indirectement spécifie l'action «venir à leur rencontre» en (36). L'hypéronymie se situe à deux niveaux de hiérarchie différents : «répondre» est un hyponyme de «dire» et «bafouiller» est un hyponyme de «dire» mais bien évidemment «répondre» et «bafouiller» ne sont pas des co-hyponymes. La relation d'hyperonymie s'établit à deux niveaux différents qui peuvent, en quelque sorte, se distribuer l'un sur l'autre, Cf. Figure 7.1.

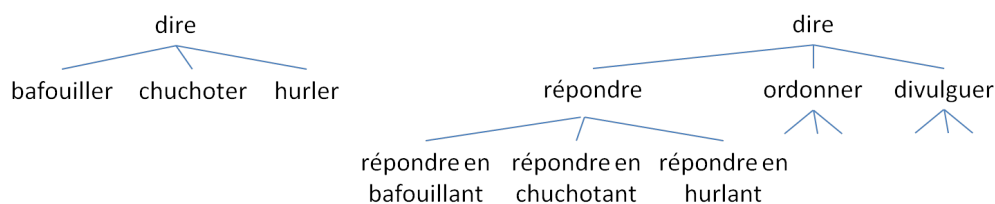


FIG. 7.1 – Deux hiérarchies d'hyperonymie dont l'une se distribue sur l'autre

Seules les configurations particulières avec le verbe «dire» sont des cas de redondance. Dans les autres configurations, le contenu sémantique du verbe régissant est plus riche, apportant des informations qui ne sont pas présentes dans le contenu sémantique du gérondif. On ne peut parler que de redondance partielle.

À présent que nous avons décrit l'éventail des interprétations possibles des constructions gérondives, nous allons proposer notre analyse de ces constructions au moyen des relations de discours en nous focalisant sur les constructions que nous traitons au moyen de la relation d'*Élaboration*. Notre analyse *via* les relations de discours va entraîner un redécoupage de la classification de Halmøy, dont nous proposons un premier aperçu dans le tableau page suivante.

Configuration	Exemple	Relation de discours
RT	<ul style="list-style-type: none"> ● En approchant de la chambre, j'entendis les cris douloureux d'un homme malgré la porte fermée. ● En rentrant dans la chambre, il se rendit compte qu'elle n'avait été tirée d'embarras que pour un embarras plus grand encore. 	<i>Arrière-plan</i> <i>Narration</i>
A <effet, cause>	<ul style="list-style-type: none"> ● Marion a pleuré en m'entendant crier. ● Alors, en ne signant pas, vous retardez la sépulture de votre fils d'autant. 	<i>Explication</i> <i>Résultat</i>
A <conséquence, condition>	<ul style="list-style-type: none"> ● Il se porterait mieux en mangeant moins. 	<i>Condition</i>
A <résultat moyen>	<ul style="list-style-type: none"> ● En prêtant l'oreille, je parvenais à discerner quelques bribes de leur conversation. ● Il essayait de se réchauffer les pieds en les frottant contre ses mains. ● George W. Bush espère éviter la récession en baissant les impôts 	<i>Résultat</i> <i>Élaboration</i> <i>But</i>
A'	<ul style="list-style-type: none"> ● Ségolène Royal avait volé dans les plumes de ses collègues en annonçant une réforme de l'accouchement sous X. 	<i>Élaboration</i>
B	<ul style="list-style-type: none"> ● À ses côtés, Jacques Chirac envoyait des baisers à la foule tout en aplatisant une mèche folle, dérangée par la brise. 	??
B'	<ul style="list-style-type: none"> ● Elle répondit en bafouillant que c'était la fille d'une ancienne voisine. 	<i>Élaboration</i>

7.3 Représentation discursive du gérondif

La variété d'interprétation possible avec cette construction justifie une analyse sémantique. Nous allons maintenant analyser les configurations de Halmøy avec des outils différents. L'analyse d'Halmøy en termes de relations temporelles et/ou causales nous a naturellement amenée à proposer, dans le prolongement de son analyse, une analyse en relations de discours (dont les effets, on l'a vu, s'expriment en termes de relations sémantiques temporelles et/ou causales). Nous allons nous concentrer sur les configurations où on a potentiellement *Élaboration* : A, A', B et B'. Nous écartons de notre analyse les syntagmes gérondifs qui sont des repères temporels, pour lesquels la relation d'*Élaboration* ne sera jamais inférée. Parmi les syntagmes gérondifs de la configuration A, nous écartons les configurations <effet, cause> et <conséquence, condition> pour lesquels une relation causale ou de condition est toujours inférée.

Pour les configurations A <résultat, moyen> et A', notre analyse en relations de discours, s'appuyant sur la distinction entre éventualité causative complexe et prédicat de critère (Sæbø, 2006), nous amène à proposer un découpage différent.

7.3.1 Configuration A <résultat, moyen>

La configuration A <résultat, moyen> correspond aux emplois typiques de la construction gérondive de la même façon que pour le marqueur *indem* étudié par Fabricius-Hansen

et Behrens (2001).

(39) Das Leni selbst eine «Partikularistin» ist, beweist sie täglich, **indem** sie sämtliche Brötchenkrümel vom Teller aufliest und in den Mund steckt (Fabricius-Hansen et Behrens, 2001 : 6)

(39)' That Leni herself is a «Particlist» is something she proves daily **by gathering up** all the bread crumbs from her plate and putting them into her mouth. (Traduction OMC)

(39)" Le fait qu'elle se considère elle-même comme une «Particulariste», Leni le prouve chaque jour **en ramassant** toutes les miettes de pain de son assiette pour les porter à sa bouche. (Notre traduction)

Fabricius-Hansen et Behrens (2001) analysent ce type d'exemple au moyen de la relation d'*Élaboration* entre le verbe régissant et la proposition introduite par *indem*. Elles s'appuient sur le sémantisme du verbe régissant «beweisen» («prouver») qui dénote une éventualité causative complexe, i.e. une éventualité complexe structurée causalement avec un résultat spécifique. Elles s'appuient sur la proposition de Kamp et Roßdeutscher (1994) pour le traitement dans le cadre de la DRT de ce type d'éventualité, contenant implicitement et de façon non spécifiée le sous-événement qui est à l'origine du résultat, Cf. Figure 7.2.

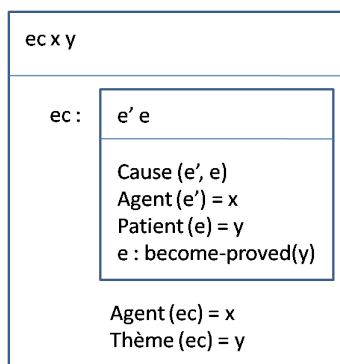


FIG. 7.2 – DRS pour «she proves something»

Cette représentation introduit deux sous-éventualités implicites e et e' d'une éventualité complexe ec «Leni prouve qu'elle se considère comme une Particulariste» et deux individus x (référent x rattaché au nom propre Leni) et y (référent d'état) impliqués dans cette éventualité. La variable y représente la proposition «Leni se considère comme une 'Particulariste'», objet (patient) du verbe «prouver». e représente le changement que la proposition y subit : le changement entre ne pas avoir la propriété d'être prouvé à celle de l'être. Enfin e' représente l'éventualité à l'origine de ce changement. L'agent de e' et de e doit nécessairement être le même. La sous-éventualité causative e' est non seulement implicite, mais aussi sous-spécifiée. La sous-éventualité e est, en revanche, donnée avec la description de l'éventualité complexe.

Dans la suite du traitement, la proposition introduite par *indem* «sie sämtliche Brötchenkrümel vom Teller aufliest und in den Mund steckt» vient spécifier la sous-éventualité

e' . Fabricius-Hansen et Behrens (2001) propose donc que *indem* relie les deux propositions par la relation d'*Élaboration*, suivant la définition donnée dans Asher et Lascarides (2003), à condition d'omettre la condition de complexité. La proposition introduite par *indem* introduit un référent d'éventualité qui s'identifie au référent e' de l'éventualité causative.

Kleiber et Vassiliadiou (2009) comparent ce type de configuration intraphrastique à des configurations où la même relation apparaît entre deux phrases indépendantes comme en (40) :

(40) Paul a éteint le feu. Il a pissé dessus. (Kleiber et Vassiliadiou, 2009, p. 200)

Dans ce cas, (40) peut être équivalente à la configuration suivante avec le gérondif «en pissant dessus» :

(41) Paul a éteint le feu **en pissant** dessus. (Halmøy, 1982, p. 231)

Comme ils le soulignent, la seconde phrase, en (40), explicite le moyen employé par Paul pour éteindre le feu. Suivant les considérations de Fabricius-Hansen et Behrens (2001), on serait donc en droit de parler d'*Élaboration* pour traiter l'exemple (40) en considérant que l'éventualité «éteindre le feu» est une éventualité causative complexe contenant deux sous-événements avec e le changement d'état et e' l'éventualité à l'origine de ce changement. La deuxième phrase, en (40), vient alors spécifier cette e' .

Kleiber et Vassiliadiou (2009) émettent deux critiques à cette analyse, réfutant l'inférence d'*Élaboration* pour cet exemple.

Ils notent que le prédicat *Subtype_D* ne peut pas être inféré en (40) entre «éteindre le feu» et «pissier sur le feu». Cependant, cet argument n'est pas suffisant car toutes les élaborations ne sont pas inférées *via* ce prédicat. Ils notent ensuite que «pissier sur le feu» n'est pas une partie d'«éteindre le feu» mais que la durée des deux éventualités semble plutôt coïncider. De notre point de vue, cet argument n'est pas non plus suffisant pour rejeter *Élaboration*, simplement parce que les effets sémantiques d'*Élaboration* autorise la coïncidence temporelle. La deuxième critique de Kleiber et Vassiliadiou (2009) porte sur le fait que ce traitement va s'appliquer à des cas plus larges, généralement analysés au moyen de la relation d'*Explication*.

(42) Paul est tombé. Il a trébuché.

(43) Paul est tombé en trébuchant.

Dans cet exemple canonique de la relation d'*Explication*, «tomber» peut effectivement être analysé comme une éventualité causative complexe contenant deux sous-événements avec e le changement d'état («d'être debout» à «être par terre») et e' l'éventualité à l'origine de ce changement («trébucher»).

Nous nous accordons sur ce dernier argument. La relation d'*Élaboration* ne nous semble pas être la relation reliant les deux propositions en (40) tout comme en (39). Nous proposons la relation de *Résultat*. En (40), «éteindre le feu» présente le résultat et «pissier sur le feu» le moyen employé pour atteindre ce résultat.

Pour les constructions gérondives de la configuration A <résultat, moyen>, une seule relation de discours ne nous paraît pas suffisante pour rendre compte de la variété discursive exprimée par le moyen. La relation de discours va nous semble-t-il dépendre de la façon dont le procès du verbe régissant est décrit. Le gérondif peut exprimer le moyen employé dans un procès en cours de déroulement (moyen₁), le moyen ayant permis d'obtenir un résultat ou le moyen employé pour atteindre un but (moyen₂).

7.3.1.1 Moyen₁

Parmi les gérondifs de la configuration A <résultat, moyen>, certains syntagmes gérondifs expriment le **moyen employé dans un procès en cours de déroulement**. Nous reprenons, à titre d'illustration, l'exemple (16).

(16) Il essayait de se réchauffer les pieds **en les frottant** contre ses mains.

En (16), le verbe «essayer», en tant que verbe d'effort, introduit le procès «se réchauffer les pieds» comme étant en cours de déroulement. De ce fait, l'événement «en frottant ses pieds contre les mains» décrit une action participant à ce procès, qui est une élaboration de la façon dont le protagoniste essaie de se réchauffer les pieds. Dans cet exemple, la relation inférée est *Élaboration*.

Il n'y a pas de subsumption possible entre l'éventualité dénotée par la proposition régissante et celle dénotée par le syntagme gérondif. En (16), on peut, éventuellement, considérer que tout événement, comme «frotter les pieds contre ses mains», peut constituer un «essai de se réchauffer les pieds», bien que les limites de cette analyse se fassent vite ressentir, comme en (44).

(44) Il essayait de se réchauffer les pieds en chantant.

Une interprétation tirée par les cheveux permet de comprendre «en chantant» comme un moyen (et non pas comme un sous-type) employé pour se réchauffer les pieds. En revanche, si le chant est considéré comme une tentative de diversion, qui n'est pas sans nous rappeler les longues montées en télésiège dans le froid, l'interprétation sera alors de considérer «en chantant» comme une circonstance accompagnante.

Pour rendre compte des gérondifs exprimant le moyen, nous ne pouvons pas faire appel au prédicat *Subtype_D*. L'interprétation va dépendre de la capacité à considérer le syntagme gérondif comme un moyen d'accomplir le procès dénoté dans le verbe régissant à partir de connaissances du monde et du contexte situationnel (i.e. du discours), ce qui pourrait être encodé dans un prédicat *Moyen_D*. Pour inférer, ici, *Élaboration*, il faut donc reconnaître en α un prédicat abstrait présentant un procès dans le cours de son déroulement et en β un syntagme gérondif pouvant, sur la base de connaissances du monde, exprimer un moyen d'accomplir ce procès.

Axiome 7.1 Inférer *Élaboration Moyen*

$$(?(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge Événement_{en\ cours\ de\ déroulement}[\alpha] \wedge Moyen_D(\sigma, \beta, \alpha) > Élaboration(\alpha, \beta, \lambda))$$

7.3.1.2 Moyen₂

Lorsque dans les constructions gérondives de la configuration A <résultat, moyen>, le syntagme gérondif exprime un résultat, une relation de causalité est exprimée entre les deux propositions de la construction. Nous reprenons, à titre d'illustration, l'exemple (17) :

(17) **En prêtant** l'oreille, je parvenais à discerner quelques bribes de leur conversation.

En (17), le verbe régissant décrit le résultat («parvenir») d'un procès («discerner quelques bribes de leur conversation»). Le procès «en prêtant l'oreille» décrit le moyen permettant d'obtenir ce résultat. L'exemple suivant présente la même configuration :

(45) **En redessinant** son réseau commercial et **en proposant** des crédits à taux zéro, B&O a réussi à faire franchir le seuil de ses magasins à des consommateurs auparavant effrayés par les prix des produits. (Halmøy, 2003, p. 97)

Les verbes «parvenir» et «réussir», en tant que verbes d'aboutissement, introduisent respectivement «discerner quelques bribes de leur conversation» et «faire franchir le seuil de ses magasins» comme des résultats. Nous inférons la relation de *Résultat*.

Dans tous les cas, le résultat est exprimé par la proposition régissante et la cause (ou le moyen) par le syntagme gérondif, quel que soit l'ordre dans lequel ils apparaissent. En suivant ainsi l'ordre dans lequel les constituants sont introduits pour donner une représentation de la structure du discours, comme le postule la SDRT pour les propositions indépendantes, nous avons deux constituants α et β , dans l'ordre dans lequel ils apparaissent dans le texte. Nous inférons $Cause_D(\alpha, \beta)$ dans les cas où la cause (dénotée par le gérondif) est exprimée avant le résultat (dénoté par le verbe régissant) et $Cause_D(\beta, \alpha)$ dans le cas contraire. Ces informations permettraient alors d'inférer soit *Résultat*, soit *Explication*. Or nous envisageons, dans les cas de subordination syntaxique, de tenir compte des dépendances syntaxiques entre les deux propositions et de proposer que la proposition principale soit toujours décrite dans le constituant β (donnant le résultat) et le syntagme gérondif dans le constituant α . Ces informations permettraient toujours d'inférer *Résultat*.

Dans la même catégorie, nous classons les syntagmes gérondifs qui expriment des moyens pour atteindre un but comme illustré en (46) :

(46) George W. Bush espère éviter la récession en baissant les impôts. (Halmøy, 2003, p. 97)

La relation ici inférée est celle de *But*.

7.3.1.3 Bilan

Ces deux interprétations possibles de la configuration A <résultat, moyen> reposent sur l'emploi d'un verbe d'effort comme *essayer* pour l'*Élaboration*, d'aboutissement comme *réussir* pour le *Résultat* ou d'espoir comme *espérer* pour le *But*. Un autre indice pouvant permettre de distinguer *moyen₁* et *moyen₂* est la télicité du verbe.

(47) Il a soigné sa blessure **en passant** une pommade aux huiles essentielles.

(48) Il a guéri sa blessure **en passant** une pommade aux huiles essentielles.

Le verbe *soigner* est atélique («il a soigné sa blessure pendant un mois»). Il présente une action dans son cours de déroulement, comme inachevée. Le verbe *guérir* est télique («il a guéri sa blessure en un mois»). Il présente une action comme menée à son terme. Suivant notre proposition, nous inférons *Élaboration* en (47) et *Résultat* en (48).

7.3.2 Configuration A'

Sæbø (2006) analyse ce qu'il appelle *les prédicats abstraits*, qui rassemblent d'un côté les éventualités causatives complexes et les prédicats de critère. Les prédicats de critère sont, en quelque sorte, sous-spécifiés. Une information supplémentaire est nécessaire pour comprendre la façon dont le procès a été réalisé. Si nous prenons, par exemple, le prédicat de critère «rendre un service» : si A demande à B de lui rendre un service, B va demander de quoi il s'agit.

Les prédicats de critère sont considérés par Sæbø (2006) non seulement comme des prédicats d'événement, mais aussi comme des prédicats de prédicats d'événement, i.e. des prédicats d'événements de second ordre ayant pour arguments le prédicat abstrait et le prédicat spécifiant la façon dont le premier a été réalisé : $pc(e, e')$. Pour le prédicat de critère «rendre un service», on a e = «rendre un service» et e' = la façon dont le service a été réalisé. e' est une éventualité implicite et sous-spécifiée. Lorsque le prédicat abstrait a pour argument une proposition en *by+ing* ou un gérondif, ces derniers viennent remplir le contenu de cet e' sous-spécifié.

(49) She kept a promise **by dancing**. (Sæbø, 2006, p. 136)

(50) Il m'a rendu un service **en me déposant** devant l'université.

Sæbø (2006) propose une analyse par unification dans le cadre de la DRT que nous ne détaillerons pas ici. Comme nous l'avons signalé, les constructions gérondives des configurations A et A' imposent une relation orientée, qui se manifeste par l'impossibilité d'inverser les arguments de la construction. Le traitement proposé par Sæbø (2006) rend compte de cette asymétrie. Le gérondif peut constituer un argument du prédicat abstrait mais pas l'inverse.

De notre point de vue, la relation d'*Élaboration* est inférée en présence d'un prédicat de critère et d'un gérondif venant spécifier la façon dont ce prédicat a été réalisé, comme illustré en (23).

(23) Ségolène Royal avait volé dans les plumes de ses collègues **en annonçant** une réforme de l'accouchement sous X.

La relation est toujours orientée.

(51) Il a commis une erreur **en divorçant**.

(52) ? Il a divorcé **en commettant** une erreur.

La relation d'inclusion en (51) peut se paraphraser en «divorcer, c'était commettre une erreur». En (52), la relation d'inclusion prend une autre interprétation : il n'est pas possible de considérer que «commettre une erreur» est une façon de «divorcer». On comprend plutôt que les deux procès se déroulent de façon concomitante.

La relation s'établit entre une éventualité abstraite dénotée par la proposition régissante et une réalisation concrète dénotée par le syntagme gérondif. Le gérondif joue un rôle crucial dans la mise en œuvre de l'interprétation, comme l'attestent les deux exemples suivants :

- (53) **En tuant** sa mère, **en étranglant** le sordide, il a aussi assassiné le rêve. (Halmøy, 2003, p. 100)
- (54) Il a tué sa mère. Il a étranglé le sordide. Il a aussi assassiné le rêve.

Dans la construction gérondive, on comprend que «tuer sa mère» et «étrangler le sordide» constituent des réalisations concrètes de l'éventualité métaphorique «assassiner le rêve». Halmøy (2003) précise que le marqueur *aussi* dans la construction gérondive oriente l'interprétation de cet exemple vers une relation d'équivalence entre les deux procès alors que dans la construction où les trois propositions sont juxtaposées, *aussi* oriente plutôt l'interprétation vers une succession d'éventualités. Incontestablement, le gérondif joue un rôle primordial dans l'interprétation de la relation d'*Élaboration* en indiquant que le syntagme gérondif entretient un lien étroit avec le verbe régissant.

Dans tous les cas, le constituant introduit par le syntagme gérondif est le constituant élaborant et le constituant introduit par le verbe régissant est le constituant élaboré. Lorsque le syntagme gérondif est en tête de la construction, le constituant élaborant est introduit avant le constituant élaboré dans une sorte d'*Élaboration* inverse. Cet ordre ne correspond pas à l'usage protoyptique de la relation et a été très peu étudié.

- (55) **En divorçant**, il a commis une erreur.

7.3.3 Configuration B

Reprenons, ci-dessous, un exemple de la configuration B :

- (25) Le chef faisait les cents pas **en fumant** sa pipe en bambou.

Pour l'analyse en relations de discours des gérondifs de la configuration B, si l'on s'appuie sur la définition de Halliday (1985) : «*The secondary clause does not introduce a new element into the picture but rather provides a further characterization on what is already there, restating it, clarifying it, refining it, or adding a descriptive attribute or comment.*», il apparaît possible de relier les deux propositions par la relation d'*Élaboration*⁶. C'est également l'analyse proposée par Kleiber (2007) sur un autre exemple de cette configuration :

⁶Lors de la conférence MAD'10 *Signalling Text Organisation*, nous avons présenté ce type d'exemples comme des cas d'*Arrière-plan*. Bateman, lors de son intervention, a plutôt proposé de relier les deux propositions par *Élaboration*.

(56) Marie est sortie en pleurant. (Kleiber, 2007, p. 118)

Le syntagme gérondif apporte des informations supplémentaires à l'éventualité décrite par le verbe régissant. Il introduit des éléments qui concourent à ajouter des informations sur ce qui est déjà présent dans la première éventualité.

Les exemples tels que (25) ou (56) ne se laissent pas facilement analyser avec les outils disponibles de la SDRT. Le syntagme gérondif n'élabore pas, dans le sens classique des cas que nous avons déjà analysés, l'éventualité dénotée par la proposition principale mais définit des circonstances dans lesquels celle-ci a eu lieu. Mais la relation d'*Arrière-plan* nous semble également exclue. Le gérondif est neutre du point de vue aspectuel et le syntagme gérondif ne semble pas définir un arrière-plan à l'éventualité décrite dans la proposition principale.

Il nous semble que les deux éventualités décrites dans ces constructions gérondives participent à la description d'une éventualité complexe, que l'on peut reconstruire par la somme du contenu des deux éventualités. Mais, comme le souligne (Kleiber, 2007), les deux procès ne sont pas simplement associés sur un même niveau, le gérondif implique une intégration d'une prédication dans l'autre.

Les constructions gérondives de la configuration B décrivent deux procès, dont la relation est fortuite, se déroulant de façon concomitante, dans le même endroit et impliquant un même agent. Behrens et Fabricius-Hansen (2005) introduisent une nouvelle relation *Circonstance accompagnante* (*Accompanying circumstance*) pour rendre compte de ces cas. Cette relation, vérifiée entre deux propositions, est empruntée à Kortmann (1991). Elle peut être définie de façon informelle comme une relation s'établissant entre deux propositions dénotant deux états ou deux événements qui «from the point of view of the speaker/writer, form a unit mentally» (Kortmann, 1991, p. 122). Les deux propositions forment ensemble une éventualité complexe.

Elles situent cette nouvelle relation dans l'ensemble des relations de la SDRT. Elles signalent que, tout comme la relation d'*Élaboration*, la relation de *Circonstance Accompagnante* est vérifiée entre deux éventualités (deux états ou deux événements) si celles-ci ont lieu en même temps, au même endroit et impliquent un même agent. En revanche, lorsque les deux éventualités sont reliées par *Circonstance accompagnante*, le syntagme gérondif ne décrit pas la même action que celle décrite par la proposition principale, mais plutôt deux facettes d'une action.

Behrens et Fabricius-Hansen (2005) proposent que les deux éventualités reliées par la relation de *Circonstance accompagnante* donnent lieu à la construction d'un topique dominant dont le contenu est en quelque sorte la somme des descriptions des deux éventualités. Pour Behrens et Fabricius-Hansen (2005), cette relation serait donc coordonnante. Elles proposent donc de considérer les deux relations coordonnantes existantes afin de voir si celles-ci permettent de rendre compte des exemples tels que (25).

La relation de *Narration* est exclue lorsqu'il s'agit de rendre compte d'un exemple tel que (25). Tandis que la relation de *Narration* entraîne que les deux éventualités se succèdent temporellement, les deux éventualités en (25) co-occurrent dans le temps.

Continuation est, quant à elle, définie comme une relation coordonnante sans contrainte spatio-temporelle. Comme le soulignent Asher et Lascarides (2003), *Continuation* n'est pas une relation impliquant la somme de ses arguments. C'est une relation pour laquelle le topique est déjà en place et pour laquelle les arguments de la relation partagent une même fonction rhétorique vis-à-vis de ce topique. La relation semble, de ce fait, adéquate dans le sens où les deux éventualités reliées par *Circonstance Accompagnante* ont la même fonction rhétorique d'élaboration vis-à-vis du topique dominant et inadéquate du fait que la relation de *Continuation* n'implique pas l'opération de somme de ces deux éventualités.

En second lieu, Behrens et Fabricius-Hansen (2005) notent que *Circonstance Accompagnante* se comporte comme une relation subordonnante dans le sens où la contrainte de la frontière droite est respectée :

(57) [The others followed her,]_a [waving their weapons.]_b [Then they stopped.]_c

Le segment (c) introduit par le connecteur *then* se rattache au segment (a) par la relation de *Narration*. Cet attachement n'est possible qu'à condition que le segment (a) se trouve sur la frontière droite, i.e. que le segment (b) lui soit attaché par une relation subordonnante. Effectivement si le segment (b) est attaché au segment (a) par une relation coordonnante, celui-ci ne se trouverait plus sur la frontière droite et ne serait donc pas disponible pour de futures continuations.

Elles concluent que *Circonstance Accompagnante* est à la fois subordonnante et coordonnante. Elle est coordonnante dans le sens où elle implique la construction du topique dominant dont le contenu est la somme du contenu des arguments de la relation et elle se comporte comme une relation subordonnante vis-à-vis de la frontière droite.

D'après Behrens et Fabricius-Hansen (2005), une nouvelle relation de discours, *Circonstance Accompagnante*, devrait être introduite dans le cadre de la SDRT pour pouvoir rendre compte de ces cas, mais, de notre point de vue, cette relation ne peut être à la fois coordonnante et subordonnante.

Comme nous allons le voir, le traitement que nous proposons des constructions gérondives de la configuration B reprend celui qui est présenté ci-dessus. Il nous semble effectivement qu'il est nécessaire d'ajouter une nouvelle relation de discours au cadre de la SDRT, la relation de *Circonstance accompagnante* pour conserver la terminologie proposée par Behrens et Fabricius-Hansen (2005).

Contrairement à Behrens et Fabricius-Hansen (2005), nous ne considérons pas la double nature coordonnante et subordonnante de la relation.

Solution 1 - Relation coordonnante

Dans cette solution, la relation de *Circonstance Accompagnante* est considérée comme coordonnante, inférée entre un constituant α obtenu à partir de la proposition principale et un constituant β obtenu à partir du syntagme gérondif. Nous proposons que cette relation ait les effets sémantiques suivants :

- des conséquences temporelles :

Axiome 7.2 *Circonstance Accompagnante* – *Conséquence temporelle*

$$\phi_{\text{Circonstance Accompagnante}(\alpha, \beta)} \Rightarrow \alpha \equiv_t \beta$$

- une contrainte de topique :

Axiome 7.3 *Circonstance Accompagnante* – *Contrainte Topique*

$$\phi_{\text{Circonstance Accompagnante}(\alpha, \beta)} \Rightarrow \neg \Box(\mathcal{K}_\alpha \sqcap \mathcal{K}_\beta)$$

Cet axiome permet la création d'un topique implicite au moyen de l'opérateur \sqcap qui calcule le contenu commun des constituants. L'opérateur modal de nécessité \Box spécifie que ce topique ne peut pas être vide de sens.

- une contrainte de topique au moment de la mise à jour :

Axiome 7.4 *Circonstance Accompagnante* – *Contrainte Topique MAJ*

$$\text{Circonstance Accompagnante}(\alpha, \beta, \lambda) \rightarrow \exists \delta((\delta = \alpha \sqcap \beta) \wedge \exists \gamma(\text{Élaboration}(\gamma, \lambda, \delta)))$$

Cet axiome, intervenant au moment de la mise à jour de la SDRS, permet la création d'un constituant complexe formé des deux étiquettes reliées par *Circonstance Accompagnante*. Ce constituant est alors dominé par le constituant implicite au moyen de la relation d'*Élaboration*.

L'apparent problème de la frontière droite exposé par Behrens et Fabricius-Hansen (2005) est résolu avec la création du topique. En effet, en (57), (c) sera directement attaché au constituant topique (π^*), tel que représenté dans la Figure 7.3.

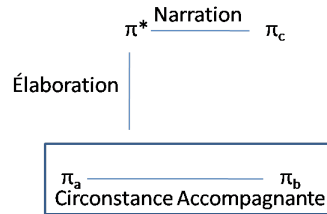


FIG. 7.3 – Représentation de (57)

Avec cette solution, les deux propositions de la construction gérondive B sont situées au même niveau, reliées par une relation coordonnante. De plus, cette solution implique la construction d'un topique implicite dominant les deux propositions par *Élaboration*. Mais elle ne rend pas compte du fait que le contenu du syntagme gérondiv vient enrichir le contenu de la proposition principale et pas l'inverse.

Solution 2 - Relation subordonnante

Dans cette solution, la relation de *Circonstance Accompagnante* est considérée comme subordonnante, rendant plus efficacement compte du rôle de la procédure d'intégration de la prédication du syntagme gérondif dans celle de la proposition principale. Cette relation pourrait être considérée comme un sous-type de la relation d'*Élaboration* avec des effets sémantiques plus restreints (on passe de l'inclusion à l'équivalence) :

Axiome 7.5 *Circonstance Accompagnante* _ *Conséquence temporelle*

$$\phi_{\text{CirconstanceAccompagnante}(\alpha, \beta)} \Rightarrow \alpha \equiv_t \beta$$

Enfin, pour rendre compte de l'idée que les deux propositions décrivent deux facettes d'une même action, ayant poussé Behrens et Fabricius-Hansen (2005) à la considérer d'une certaine façon comme coordonnante, il faut que le topique soit mis à jour avec les informations contenues dans le constituant dominé.

Enfin, nous considérons le traitement proposé pour les constructions gérondives de la configuration B'.

7.3.4 Configuration B'

Suivant les descriptions d'Halmøy (2003), les constructions gérondives de la configuration B' décrivent une seule action. Elle distingue deux cas typiques dans lesquels le verbe régissant est soit un verbe de parole prototypique «dire» ou un verbe de déplacement prototypique «se déplacer».

(58) [dit-il]_a [en bafouillant]_b \Rightarrow bafouilla-t-il (Halmøy, 2003, p. 105)

(59) [Il se déplaça]_a [en faisant des petits pas.]_b

Halmøy (2003) décrit la relation entre le gérondif et le verbe régissant comme une relation d'hyponymie. Dans le cadre de la SDRT, nous pouvons dans les deux cas ci-dessus inférer le prédicat *Subtype_D*. Le type de «bafouiller» est un sous-type de celui de «dire» : «bafouiller» implique «dire» mais pas l'inverse. De même le type de «faire des petits pas» est un sous-type de celui de «se déplacer» : «faire des petits pas» implique «se déplacer» mais pas l'inverse. En s'appuyant sur l'inférence du prédicat *Subtype_D*, il est possible d'inférer *Élaboration* entre les segments (a) et (b) en (58) et (59).

La situation est plus complexe dans les exemples (35) et (36) également traités chez Halmøy (2003) comme cas de la configuration B' dont nous reprenons ci-dessous deux exemples qui ont déjà été discutés dans la section 7.2.5 :

(35) Elle répondit **en bafouillant** que c'était la fille d'une ancienne voisine.

(36) Ludi vient à leur rencontre **en galopant**.

En effet, on ne peut pas considérer le type de «bafouiller» comme un sous-type de «répondre» : «bafouiller» n'implique pas «répondre». De la même façon, on ne peut pas considérer le type de «galoper» comme un sous-type de «venir à leur rencontre» : «galoper» n'implique pas «venir à leur rencontre». En revanche, dans les deux cas on peut considérer que la relation de sous-type s'applique de façon indirecte étant donné que «répondre» induit un acte de parole et «venir à leur rencontre» induit un déplacement. De ce fait, pour représenter au mieux la relation s'établissant entre le verbe régissant et le gérondif, qui apporte des indications sur l'acte de parole ou l'acte de déplacement, nous pensons juste d'inférer la relation d'*Élaboration*.

7.4 Conclusion

La valeur interprétative de la relation liant le syntagme gérondif et la proposition régissante naît de la relation existant entre le sémantisme des verbes et des connaissances du monde. Nous tenons également compte de la façon dont le procès de la proposition régissante est présenté (dans son déroulement ou son résultat). Toutes ces informations nous permettent de considérer la construction gérondive comme un indice de la relation d'*Élaboration*.

Configuration	Éléments responsables de l'inférence d' <i>Élaboration</i>	Schéma (avec <i>a</i> pour la proposition principale et <i>b</i> le syntagme gérondif)
A <résultat, moyen>	<p>verbe d'aboutissement : <i>réussir</i>, <i>parvenir</i> ou éventualité causative complexe</p> <p>verbe d'espoir : <i>espérer</i>, <i>chercher</i></p> <p>verbe d'effort : <i>essayer</i>, <i>s'efforcer de</i>, <i>tenter</i></p>	
A'	prédicat de critère : <i>voler dans les plumes</i>	
B	deux éventualités concomitantes	
B'	hypéronymie d'un verbe de dire ou de déplacement	

La construction gérondive n'étant pas un marqueur d'*Élaboration*, nous proposerons, dans le chapitre 10, pour poursuivre cette analyse, de combiner le marqueur du syntagme gérondif à des marqueurs relevant de la lexicale afin de distinguer des constructions gérondives qui sont des cas d'*Élaboration* de celles qui ne le sont pas.

Dans ce chapitre sur la construction gérondive, nous avons uniquement regardé les relations intraphrastiques. Nous avons vu que l'interprétation d'un syntagme gérondif doit donner lieu à une interprétation cohérente. Le lecteur cherche, en quelque sorte, l'interprétation la plus accessible : celle de moyen, de manière, de cause, de condition... Kleiber et Theissen (2006) soulignent que peu de travaux ont regardé les relations qu'entretient la construction gérondive avec le reste du texte. Comme ils le font pour les gérondifs RT, nous envisageons de regarder les contextes des constructions gérondives comme illustré en (60) :

(60) Tout le monde s'est essayé à faire deux choses à la fois, comme conduire en parlant.

Troisième partie

Approche en corpus de l'*Élaboration*

Deux approches permettent d'aborder la question de la signalisation d'une relation de discours.

L'approche sémasiologique part d'un marqueur potentiel pour en appréhender les fonctionnements dans la mise en œuvre d'une relation de discours. C'est ainsi que nous avons procédé pour nos travaux présentés dans la partie 2. En adoptant cette approche, il est assez facile d'interroger un corpus pour en extraire les concordances des marqueurs potentiels.

L'approche onomasiologique part de la relation de discours pour découvrir la façon dont celle-ci est signalée dans les textes. Pour adopter cette approche de façon systématique, il est nécessaire d'avoir un corpus annoté en relations de discours. Nous avons contribué à la contribution d'un tel corpus, le corpus ANNODIS, unique en ce qui concerne le français. Pour étudier une relation de discours comme celle d'*Élaboration*, généralement décrite comme une relation qui n'a pas de marqueur spécifique, disposer d'un tel corpus permet d'accéder aux réalisations linguistiques de celle-ci.

Dans cette partie 3, nous nous situons plus largement dans une perspective de linguistique de corpus. Dans la partie 2, nous nous plaçons dans une perspective d'analyse des marqueurs du discours, i.e. nous cherchions à déterminer le rôle et le fonctionnement discursif d'un marqueur dans la mise en œuvre d'une relation de discours. Dans la partie 3, nous sommes à la recherche des indices et configurations d'indices favorables à la présence d'une structure élaborative. Nous nous appuyons, pour cet objectif, sur les données du corpus ANNODIS présenté dans le chapitre 8. L'objectif n'est plus de déterminer si les éléments considérés jouent un rôle dans la mise en œuvre de la relation de discours mais plus largement si leur présence est corrélée avec celle d'une structure élaborative (chapitre 9). Dans cette optique, nous cherchons également à dégager des patrons pour l'identification automatique de la relation d'*Élaboration* (chapitre 10).

Chapitre 8

Présentation du corpus ANNODIS

Sommaire

8.1	Préparation de l'annotation	236
8.1.1	Objets à annoter	236
8.1.1.1	Perspective descendante	236
8.1.1.2	Perspective ascendante	238
8.1.2	Annotation exploratoire	239
8.2	Déroulement de l'annotation et manuels	239
8.2.1	Perspective descendante	239
8.2.2	Perspective ascendante	240
8.2.2.1	Spécifications pour la segmentation en UDE	240
8.2.2.2	Spécifications pour l'annotation en relations de discours	242
8.2.2.3	Liste de marqueurs potentiels	251
8.2.3	Outil d'annotation	252
8.2.4	Exploitation des données	252
8.3	Perspectives	252

Le projet ANNODIS¹ (2007-2010) arrive à son terme en même temps que notre travail de thèse. L'objectif de ce projet, auquel nous avons participé, est de fournir un corpus annoté avec des informations discursives et des outils informatiques permettant son exploitation (Péry-Woodley *et al.*, 2009)².

Ce corpus propose deux types d'annotations.

L'approche dite *descendante* aborde le texte dans son entier à la recherche de structures de haut niveau jouant un rôle dans l'organisation du texte.

L'approche dite *ascendante* part des plus petits segments de texte pour atteindre la globalité du texte *via* des relations de discours.

¹Le projet ANNODIS a été financé par l'ANR - Programme Sciences Humaines et Sociales Appel 2007 «Corpus et outils de la recherche en sciences humaines et sociales». Ce projet, dirigé par Marie-Paule Péry-Woodley, réunit trois laboratoires de recherche : CLLE-ERSS, Toulouse ; IRIT, Toulouse et GREYC, Caen.

²Site web du projet : <http://w3.erss.univ-tlse2.fr/textes/pagespersos/annodis/index.html>

Adopter l'une ou l'autre de ces deux démarches implique une façon différente d'aborder le texte. L'approche ascendante, qui cherche à relier un à un tous les segments de texte *via* des relations de discours, considère des textes courts ou des extraits de textes longs (50 à 1000 mots). L'approche descendante, qui cherche à identifier des structures de haut niveau participant à l'organisation du texte, considère des textes entiers longs et structurés (avec des sections et sous-sections) (6000 à 15000 mots).

Un outil, GLOZZ, a été spécialement créé pour l'annotation manuelle de ces structures discursives et l'exploitation des annotations produites.

Ce corpus présente trois caractéristiques principales :

- C'est un corpus diversifié composé d'articles du journal l'Est Républicain³ (ER) (pour les textes courts), d'articles de Wikipedia (Wiki), d'articles de géopolitique⁴ (Geopo) et d'articles scientifiques en sciences du langage⁵ (pour les textes longs).
- C'est un corpus qui sera disponible à la fin du projet, composé de documents libres de droits.
- C'est un corpus d'étude pour les perspectives ascendantes (textes courts ou extraits de ER, Wiki et CMLF) ; pour les perspectives descendantes (textes longs, découpés en plusieurs sections et sous-sections...) et pour la rencontre des deux perspectives (une dizaine d'extraits de textes longs de la perspective descendante ont été également annotés selon la perspective ascendante).

Nous proposons, dans ce chapitre, de présenter les différentes étapes de l'élaboration du corpus ANNODIS, de la préparation (section 8.1) au déroulement de l'annotation (section 8.2). Enfin, nous concluons sur les bénéfices à tirer d'un tel corpus pour nos recherches de marqueurs et indices de la relation d'*Élaboration* (section 8.3).

8.1 Préparation de l'annotation

8.1.1 Objets à annoter

8.1.1.1 Perspective descendante

Dans la perspective descendante, deux structures ont été annotées - les structures énumératives (Figure 8.1) et les segments correspondant à une chaîne référentielle (Figure 8.2) - et les indices permettant de les identifier.

Les structures énumératives, comme nous les avons décrites dans le chapitre 6, sont généralement composées de trois segments :

- une *amorce*, segment qui annonce une énumération. Cette amorce peut contenir un SN *prospectif* (ou prospect), qui est un SN lexical à valeur cataphorique. Les prospects peuvent être introduits par un numéral, un indéfini (*plusieurs, quelques...*) ou un collectif (*une foule de, une grande variété de...*) et peuvent contenir un *énumérathème* qui spécifie le critère de co-énumérabilité.

³Mis à disposition par le CNRTL.

⁴Corpus constitué par Ho-Dac (2007) dans le cadre sa thèse avec des articles mis à disposition par l'IFRI.

⁵Articles issus des actes de la 1^{ère} Conférence Mondiale de Linguistique Française (CMLF).

- une *énumération*, composée d'au moins deux items.
- et enfin une *clôture* qui clôt l'énumération. Cette clôture peut également contenir un énumérathème.

Les segments correspondant à une chaîne référentielle (ou SUR = Segment ayant une Unité Référentielle) sont des segments larges dont la majorité des propositions qui les compose ont pour objet (parlent de / sont au sujet de / sont à propos de / apportent des informations sur) un seul et même référent. L'expression de ce référent passe nécessairement par la position grammaticale sujet.

Ces exemples appellent à la prudence.	Les facteurs essentiels sont ici les suivants.	amorce
En premier lieu, la guerre a en général pour les acteurs locaux un intérêt politique de premier ordre, souvent bien plus important que celui des Etats-Unis. Leur tolérance à la souffrance est donc plus grande.	En deuxième lieu, en dépit de leur taille réduite, ces acteurs supplantent d'ordinaire les Etats-Unis dans une ressource précise : le nombre d'hommes en âge de combattre. Même s'il n'est plus l'élément déterminant de la guerre terrestre, il reste un facteur critique, notamment en ville, dans la jungle ou en montagne.	item
Troisièmement, les " locaux " disposent en général d'un avantage : ils jouent à domicile. Si les états-Unis ont constitué au fil des décennies la mémoire institutionnelle qui leur permet de maintenir leur maîtrise des espaces, les acteurs locaux ont fait un travail similaire sur leur propre pays. Ils connaissent intimement le terrain et la météo (...)		item
Quatrièmement, nombre des chefs militaires de ces Etats ou entités ont été formés dans le monde développé - [...]	Cinquièmement, l'arsenal nécessaire au combat rapproché, [...] est beaucoup moins coûteux que les armements nécessaires à la guerre dans les "espaces communs".	item

FIG. 8.1 – Exemple d'une structure énumérative annotée

Ces exemples appellent à la prudence. Les facteurs essentiels sont ici les suivants.
En premier lieu, la guerre a en général pour les acteurs locaux un intérêt politique de premier ordre, souvent bien plus important que celui des Etats-Unis. Leur tolérance à la souffrance est donc plus grande. En deuxième lieu, en dépit de leur taille réduite, ces acteurs supplantent d'ordinaire les Etats-Unis dans une ressource précise : le nombre d'hommes en âge de combattre. Même s'il n'est plus l'élément déterminant de la guerre terrestre, il reste un facteur critique, notamment en ville, dans la jungle ou en montagne. Troisièmement, les " locaux " disposent en général d'un avantage : ils jouent à domicile. Si les états-Unis ont constitué au fil des décennies la mémoire institutionnelle qui leur permet de maintenir leur maîtrise des espaces, les acteurs locaux ont fait un travail similaire sur leur propre pays. Ils connaissent intimement le terrain et la météo (...)
Quatrièmement, nombre des chefs militaires de ces Etats ou entités ont été formés dans le monde développé - [...]. Cinquièmement, l'arsenal nécessaire au combat rapproché, [...] est beaucoup moins coûteux que les armements nécessaires à la guerre dans les "espaces communs".

FIG. 8.2 – Exemple d'un segment correspondant à une chaîne référentielle annoté

8.1.1.2 Perspective ascendante

Dans la perspective ascendante, les objets à annoter sont les unités de discours élémentaires (entre crochets) et les relations les reliant les unes aux autres.

- (1) [Tonneaux à Montreux-Vieux :]₁ [trois blessés graves]₂ [Trois personnes ont été gravement blessées, dimanche soir vers 17 h 40]₃ [à la suite d'un accident spectaculaire qui s'est produit à l'entrée du village de Montreux-Vieux sur la départementale 32]₄

Résultat(1,2)
Élaboration([1-2],3)
Explication(3,4)

Pour l'anglais, un corpus annoté en relations de discours a été construit dans le cadre du projet DISCOR (*DIscourse Structure and COreference Resolution*). Le but de ce projet était de tester des hypothèses sur l'interaction entre la structure du discours et la résolution de liens anaphoriques dans un texte. Les membres du projet ont fait le choix de suivre le modèle de la SDRT pour annoter des dépêches de presse issues du *Wall Street Journal* dans lesquelles étaient déjà annotées des chaînes de co-référence.

Comme nous l'avons vu dans le chapitre 1, plusieurs théories ont proposé des descriptions d'un modèle de la cohérence *via* des relations de discours. Jusqu'à maintenant, nous avons proposé des analyses dans le modèle de la SDRT (modèle théorique connu de l'ensemble des membres du projet pour la perspective ascendante). Pour cette annotation, les membres du projet ont également adopté une démarche sémantique de la structure du discours mais ont écarté l'idée de demander aux annotateurs une annotation suivant le modèle de la SDRT. Deux raisons ont guidé ce choix : d'abord, le modèle de la SDRT est complexe et une telle annotation demanderait que les annotateurs y soit longuement formés ; ensuite, pour ce premier projet d'annotation de corpus en relations de discours, les membres du projet ANNODIS ont dans l'idée d'utiliser les résultats de l'annotation pour valider ou invalider des hypothèses théoriques posées par le modèle. Il faut donc que les annotateurs n'aient pas été trop guidés par la connaissance du modèle.

Les manuels de segmentation et d'annotation en relations de discours résultent d'une adaptation du modèle de la SDRT (qui demeure un background théorique omniprésent dans l'équipe et dans le projet). Les données recueillies vont constituer des représentations des jugements des annotateurs *naïfs*, i.e. des perceptions qui sont également accessibles aux lecteurs mais de façon tacite.

Bien évidemment, dans un tel projet d'annotation, on ne peut pas être complètement athéorique.

- Concernant les textes, il est supposé que leur forme linéaire est capable de représenter fidèlement les structures du discours sous-jacentes envisagées.
- Concernant les segments de discours, ils sont supposés ne jamais se chevaucher.
- Concernant les relations de discours, elles sont supposées pouvoir être appliquées sans difficultés par les annotateurs *naïfs*. Il est également supposé que les mêmes

relations de discours peuvent être vérifiées entre des segments simples et des segments complexes.

- Concernant l'analyse, il est supposé que l'analyse de la réception d'un texte peut procéder de manière ascendante.

8.1.2 Annotation exploratoire

Nous avons participé à une longue période du projet, appelée *annotation exploratoire*.

Notre participation, pour l'approche ascendante, a consisté à annoter des textes et à confronter nos annotations avec celles des autres membres du projet pour la constitution des deux manuels : le manuel de segmentation et le manuel d'annotation. Une première version des manuels a été soumise à deux annotateurs pour une phase exploratoire. Ces derniers ont annoté une vingtaine de textes et leurs retours ont permis d'améliorer les manuels.

Enfin, les membres de l'approche ascendante (nous inclus) ont participé à l'annotation exploratoire de l'approche ascendante et inversement les membres de l'approche descendante ont participé à l'annotation exploratoire de l'approche ascendante. Les derniers échanges liés à ces deux annotations exploratoires ont contribué à l'élaboration des versions finales des manuels présentés dans la section suivante.

8.2 Déroulement de l'annotation et manuels

L'annotation opérationnelle a été menée par six⁶ annotateurs, dits *naïfs*, tous étudiants titulaires d'une licence ou d'un master de sciences du langage mais n'ayant pas ou peu de notions sur les théories et analyses du discours.

- La partie ascendante a réuni 42 textes doublement annotés.
- La partie descendante a réuni 87 textes annotés par un annotateur.
- 18 extraits de texte ont été annotés selon les deux approches.

8.2.1 Perspective descendante

Une des spécificités du corpus ANNODIS pour la perspective descendante est le pré-marquage de divers indices relatifs aux structures discursives à annoter. Parmi ces indices, nous relevons, comme illustré en Figure 8.1, des marqueurs d'intégration linéaire (*en premier lieu, en deuxième lieu, troisièmement, quatrièmement, cinquièmement*), des noms définis (*les facteurs, les acteurs*), des pronoms anaphoriques (*il, ils*). La tâche des annotateurs est de se baser sur ces indices pré-marqués pour annoter les structures énumératives et les segments correspondant à une chaîne référentielle. L'annotateur doit ensuite valider les indices sur lesquels il s'est appuyé pour annoter la structure et signaler les autres indices l'ayant mis sur la voie de la reconnaissance de ladite structure (Cf. Figures 8.1 et 8.2).

L'annotation se déroule en une seule étape. L'annotateur parcourt globalement le texte à la recherche de zones denses en indices pré-marqués. Une fois une zone dense définie, l'annotateur doit annoter les structures discursives qu'il détecte.

⁶Trois annotateurs pour l'analyse descendante et trois annotateurs pour l'analyse ascendante.

Concernant les structures énumératives, dont nous exploiterons dans le chapitre suivant les annotations, la tâche d’annotation consiste à

- annoter, le cas échéant, les différents éléments constitutifs de la structure : l’amorce, les items, la clôture et les énumérathèmes comme «facteurs» dans l’exemple Figure 8.1) ;
- valider les indices ayant aidé à la reconnaissance de la structure (ici, les marqueurs d’intégration linéaire *en premier lieu, en deuxième lieu, troisièmement, quatrième-ment, cinquièmement*, le SN prospectif *les facteurs* et le SN *les suivants*) ;
- et regrouper les différents éléments dans un schéma correspondant à la structure énumérative (Cf. Figure 8.1).

8.2.2 Perspective ascendante

L’annotation pour la perspective ascendante se déroule en deux étapes. Chaque texte est annoté par deux annotateurs. À partir d’un texte brut, la première étape, appelée *tâche de segmentation*, consiste à découper ce texte en unités de discours élémentaires (UDES).

Une fois que les deux annotateurs ont proposé un découpage en UDES, ceux-ci se concertent pour proposer une segmentation commune à partir de laquelle sera menée la tâche d’annotation en relations de discours. Ensuite, chaque annotateur annote le texte en relations de discours.

Nous allons, dans les sous-sections suivantes, présenter et discuter les choix qui ont été faits pour les spécifications de segmentation et d’annotation des relations de discours.

8.2.2.1 Spécifications pour la segmentation en UDE

Dans le cadre du projet ANNODIS, seules les UDE ou segments simples sont annotés. L’idée principale, lors de cette tâche, est de ne pas anticiper sur la construction d’une structure. Les autres UDE ou segments complexes sont ensuite obtenus par l’annotation des relations de discours. La tâche de segmentation consiste donc à délimiter les UDES les unes à la suite des autres, les UDES ne pouvant pas se chevaucher mais pouvant être enchâssées les unes dans les autres (par exemple les segments (a), (b) et (c) de (2) sont enchâssés dans le segment (d)). Les UDES sont décrites comme les plus petites unités pouvant être liées par une relation de discours. L’UDE prototypique est une proposition verbale indépendante correspondant à la description d’un événement ou d’un état unique comme illustré en (d) (2).

- (2) [Ces pièces, [mondialement connues,]_a [donc difficilement écoulables,]_b [et d’une valeur historique inestimable,]_c avaient été repérées chez un riche amateur nippon.]_d

Les UDES (a), (b) et (c) sont des appositions, venant spécifier l’objet «pièces». Il est moins courant et peu consensuel de considérer ces appositions comme des segments de discours. Or, nous voyons dans cet exemple que ces trois appositions sont rhétoriquement structurées. Le marqueur *donc* signale une relation de résultat entre le segment (a) et le segment (b) et la coordination *et* signale une relation de continuation entre le segment (a) et le segment (c).

Dans la phrase complexe, sont segmentées les corrélatives (3), les subordonnées conditionnelles, (4), les subordonnées circonstancielles (5), les subordonnées complétives (6), les subordonnées relatives (7)⁷, les infinitives introduites par un verbe de communication et d'attitude propositionnelle (8) ou introduites par la préposition *pour* (9) :

- (3) [Plus on est de fous,]_a [plus on rit.]_b
- (4) [S'il fait beau,]_a [j'irai à la montagne demain.]_b
- (5) [Quatre détenus de la maison d'arrêt d'Avignon ont été interpellés samedi 29 décembre sur le toit de la prison]_a [alors qu'ils tentaient de s'évader.]_b
- (6) [M. Deroose assure]_a [que ce plan sera officialisé dans les jours prochains.]_b
- (7) [Il a agressé les concurrents]_a [qui bivouaquaient près de l'aéroport de Ghadamos.]_b
- (8) [Ils auraient déclaré]_a [avoir agi en état d'ivresse.]_b
- (9) [Environ trois cents nationalistes basques ont manifestés samedi 29 décembre devant les prisons de Fleury-Mérogis, de Fresnes et de la Santé,]_a [pour réclamer la libération de la cinquantaine de militants des organisations séparatistes ETA et Iparretarrak détenus dans ces établissements de la région parisienne.]_b

En revanche, sont traitées comme des UDEs uniques les clivées (10) et les infinitives sujet (11) et objet (12) (sauf celles introduites par un verbe de communication ou d'attitude propositionnelle, Cf. ci-dessus (8)) :

- (10) [C'est un figurant qui aurait joué le rôle du vendeur.]_a
- (11) [Lire des romans est son activité préférée.]_a
- (12) [Il essayait de s'enfuir.]_a

Toutes les appositions (syntagmes adjectivaux comme dans (2), participe passé, participe présent, syntagmes nominaux) et les incises (segments (a) et (b) dans (13)) sont segmentées.

- (13) [A ce rythme là, [ajoute-t-il,]_a j'aurai, [c'est certain,]_b dépassé mon objectif de ventes.]_c

Plusieurs types d'adjoints sont segmentés. Nous segmentons les noms référents à un événement ou un état (comme *recherche*, *imprudence*, *coupure de courant*, *séquestration*, *vente*, *fatigue*...) uniquement lorsque ceux-ci sont introduits par une préposition ou locution prépositionnelle spatio-temporelle, causale, concessive ou contrastive :

- (14) [M. Pitoun est retrouvé sain et sauf,]_a [après quinze jours de séquestration.]_b

Enfin, parmi les adjoints, nous segmentons les constructions participiales (participe présent et gérondif) :

- (15) [Plusieurs individus ont mis le feu à des bouteilles d'essence posées à l'entrée du centre culturel,]_a [provoquant l'explosion de la double porte vitrée.]_b

⁷Toutes les relatives sont segmentées bien que seules les relatives non restrictives jouent un rôle au niveau de la structure discursive. Mais si nous avons fait le choix de toutes les segmenter, c'est qu'il n'est pas toujours évident d'analyser la fonction restrictive vs. non restrictives des relatives dans les textes.

- (16) [Un détenu du centre pénitentiaire de Fresnes s’est suicidé dans la nuit du 1 au 2 janvier]_a [en absorbant des barbituriques]_b

Les propositions principales ou subordonnées coordonnées sont toujours segmentées (respectivement (17) et (18)) ainsi que les verbes coordonnés partageant un même sujet et un même objet (19) et les adjoints coordonnés (20) :

- (17) [Je dois rédiger le manuel]_a [et je ne trouve pas de bons exemples.]_b
 (18) [Jean est parti]_a [parce que Marie est venue]_b [et qu’il était fatigué.]_c
 (19) [Elle a mesuré]_a [et schématisé chaque pièce de la maison.]_b
 (20) [Le village était plongé dans l’obscurité]_a [en raison d’une coupure de courant]_b [et en raison de l’absence de lune.]_c

En revanche, on ne segmente pas les sujets ou objets coordonnés :

- (21) [Il n’a cessé de lutter pour les droits de l’homme et contre les discriminations.]_a

Suivant les propositions faites dans (Vieu *et al.*, 2005), nous avons fait le choix de segmenter les adverbiaux détachés dans deux cas principaux : lorsque ceux-ci sont détachés en tête de phrase (a) ou lorsque ceux-ci sont détachés en début de segment discursif (c) en (22).

- (22) [Un an plus tard,]_a [l’équipe descendait en deuxième division,]_b [et, [en 1966,]_c Henri Germain démissionna de la présidence du club.]_d

Ce choix est soutenu par les travaux sur le rôle des cadres de discours dans l’organisation des textes (Charolles, 1997) et les travaux qui ont suivis dans le cadre de la SDRT qui considèrent les cadres de discours spatio-temporels comme des introducteurs de nouveaux topiques contenant un événement sous-spécifié (*Cf.* . 3.2.2 pour plus de détails sur le traitement des cadres de discours dans la SDRT).

D’autres cas divers sont segmentés, comme les éléments détachés en tête de phrase pour lesquels la virgule peut être remplacée par les deux points sans changement de sens :

- (23) [Investissement en 2003,]_a [l’association envisage de changer les portes d’accès à la grande salle.]_b

Les signes de ponctuation sont des éléments qui signalent souvent une frontière de segment. On segmente toutes les parenthèses, les tirets qui définissent une incise, les guillemets qui correspondent à une citation ou un discours rapporté (sauf s’il s’agit d’un mot ou d’une expression) et tous les guillemets introduits par un verbe de communication ou d’attitude propositionnelle. On segmente également les structures éventuelles à l’intérieur des parenthèses, tirets ou guillemets. Le « : » est, enfin, une marque forte de frontière de segment.

8.2.2.2 Spécifications pour l’annotation en relations de discours

Une fois le texte segmenté en UDE, l’annotation consiste à relier ces unités entre elles *via* des relations de discours. Nous distinguons *le segment cible* et *le segment à attacher*.

Le *segment à attacher* est le segment en cours de traitement et le *segment cible* le point d'attachement du segment à attacher.

Lors de l'annotation en relations de discours, la première étape consiste à déterminer le point d'attachement d'un segment, i.e. le segment cible auquel est relié le segment en cours de considération. L'attachement d'une nouvelle unité par une relation de discours peut se faire dans la même phrase, dans le même paragraphe ou dans un contexte plus large vers l'arrière dans le texte. Les propositions subordonnées, coordonnées, les appositions et les incises s'attachent systématiquement à la proposition principale.

Lorsqu'un point d'attachement est considéré, on détermine la ou les relations de discours qui lient les deux unités. Par exemple, en (24), le segment (b) s'attache au segment (a) par la relation d'*Explication*. On le note *Explication(a,b)* :

(24) [Pierre est malade.]_a [Il a trop bu hier soir.]_b

Certains segments constituent une unité avec une cohérence forte que l'on peut mettre en relation de façon globale avec un autre segment. Comme en (25), un segment complexe, (noté [b,c,d] en donnant tous les segments ou [b-d] en donnant un intervalle) s'attache au segment (a) par la relation d'*Explication*. Par ailleurs, des relations de discours s'établissent entre les segments (b), (c) et (d). On relie le segment (c) au segment (b) par la relation de *Continuation*. Le segment (d) contient le marqueur *et puis* qui indique que celui-ci s'attache au segment [b-c] par la relation de *Continuation*.

(25) [Pierre est malade.]_a [Il a mangé des huîtres avariées]_b [et il a trop bu.]_c [Et puis, il a pris froid.]

On obtient l'annotation suivante :

Explication(a,[b-d])

Continuation(b,c)

Continuation([b,c],d)

Nous allons maintenant présenter l'éventail de relations de discours qui a été retenu pour le manuel d'annotation et discuter des différences avec l'ensemble des relations de la SDRT.

Dans le projet, les relations de discours sont organisées en plusieurs groupes :

Relations de causalité

Les relations de causalité regroupent les relations d'*Explication*, de *Résultat* et de *But*.

La **relation d'Explication** lie deux segments lorsque le segment à attacher explique le segment cible de façon explicite ou non comme en (26).

(26) [L'équipe a perdu lamentablement hier.]_a [Elle avait trop de blessés.]_b

La **relation de Résultat** caractérise des liens entre deux segments portant sur deux éventualités dont le contenu du segment à attacher résulte de celui du segment cible comme en (27).

(27) [Pierre avait bu trop de vin]_a [et a donc dû rentrer chez lui en métro.]_b

La **relation de But** relie deux segments, lorsque le segment à attacher présente de façon explicite le but dans lequel l'action décrite dans le segment cible est réalisée comme en (28).

- (28) [Les chercheurs ont fait grève]_a [pour montrer leur mécontentement.]_b

Relations structurelles

Les relations structurelles regroupent les relations de *Parallèle*, de *Contraste* et de *Continuation*.

La **relation de Parallèle** porte sur deux segments ayant une construction similaire (la plupart du temps syntaxique) comme en (29).

- (29) [Jean aime Marie;]_a [il aime Claire aussi.]_b

La **relation de Contraste** relie deux segments (a) et (b) s'il y a un marqueur explicite de contraste (*mais, par contre, cependant*) comme en (30), soit par un contraste «formel» (i.e. (a) implique une proposition p et (b) implique une proposition qui est le contraire de p, avec un argument différent) comme en (31).

- (30) [Jean aime Marie,]₁ [mais il déteste Jeanne]₂.

- (31) Jean aime Marie (implique Jean aime Marie). Pierre la hait (implique avec une substitution du sujet que Pierre hait Marie, i.e. Pierre n'aime pas Marie).

La **relation de continuation** relie deux segments quand le segment à attacher continue le rôle rhétorique du segment cible, comme entre les segments (b) et (c) de (32). La relation de continuation est une relation d'appartenance à un même bloc discursif.

- (32) [Jean était fatigué]_a [parce qu'il avait beaucoup travaillé]_b [et qu'il avait peu dormi.]_c

Relations logiques

Les relations logiques regroupent les relations de *Conditionnel* et d'*Alternation*.

La **relation de Conditionnel** relie deux segments lorsque le segment cible est une hypothèse et le segment à attacher est la conséquence de cette hypothèse comme en (33).

- (33) [S'il pleut,]_a [je resterai à la maison.]_b

La **relation d'Alternation** est équivalente à la disjonction de deux constituants comme en (34).

- (34) [Soit Philippe est au bureau à l'IRIT,]_a [soit il est au Mirail.]_b

Relation de discours rapporté

Ce groupe contient une seule relation : la **relation d'Attribution**. Cette relation, non présente dans la version de la SDRT de Asher et Lascarides (2003), a été rajoutée pour les besoins de l'analyse au cours du projet DISCOR. C'est l'analyse de dépêches de presse (*Wall Street Journal*) qui a rendu nécessaire l'introduction de cette nouvelle relation⁸.

- (35) [La direction générale de Citroën a informé ses employés]_a [que les nouveaux contrats de travail prendront effet lundi prochain]_b

Relations d'exposition/narration

Les relations d'exposition/narration regroupent les relations d'*Arrière-plan*, de *Narration*, de *Flashback*, d'*Encadrement* et de *Localisation temporelle*.

La **relation d'Arrière-Plan** relie deux segments si le segment à attacher décrit une scène servant d'arrière-plan (ce segment décrit un état, correspondant souvent dans les textes narratifs à l'usage de l'imparfait) à l'événement décrit dans le segment cible comme en (36) :

- (36) [Marie entra dans la cuisine.]_a [Pierre faisait la vaisselle.]_b

Dans le cas de la relation d'*Arrière-Plan*, on prend pour convention de mettre en premier argument de la relation l'avant-plan (l'événement) et en second argument l'arrière-plan (l'état), peu importe l'ordre dans lequel les segments apparaissent dans le texte. En (36), la relation est annotée *Arrière-plan(a,b)*.

La **relation de Narration** relie deux segments qui décrivent, dans l'ordre d'apparition, deux éventualités se succédant temporellement et appartenant à la même histoire comme en (37)

- (37) [Pierre prit son parapluie.]_a [Il sortit.]_b

La **relation de Flashback** relie deux segments lorsque le segment à attacher décrit un retour en arrière dans la narration, i.e. au lieu de raconter les événements dans l'ordre dans lequel ils se sont produits, on fait un retour en arrière sur un événement qui s'est produit avant comme en (38).

- (38) [Paul a déménagé cet été.]_a [Il avait trouvé un nouvel appartement au printemps.]_b

⁸Plus exactement, deux relations *Source* et *Attribution* ont été rajouté pour l'annotation dans le projet DISCOR. À l'heure actuelle, la SDRT distingue deux relations *Source* et *Attribution*. Les deux relations sont utilisées pour relier le contenu d'un acte communicatif (donné dans le segment à attacher) à l'agent de cette acte (donné dans le segment cible). La différence entre *Attribution* et *Source* concerne les effets sur les conditions de vérité du contenu propositionnel des segments reliés. Dans le cas de *Source*, relation véridique, les deux constituants sont considérés comme «vrai» (il y a une implication de l'agent dans la véracité du discours rapporté). Dans le cas d'*Attribution*, relation non véridique, le discours rapporté n'est pas forcément considéré comme «vrai» (il y a un détachement de l'agent qui ne prend pas parti sur la véracité du discours rapporté). Dans le cadre du projet ANNODIS, tous les discours rapportés sont annotés au moyen de la relation d'*Attribution*

La **relation d'Encadrement** n'est pas une véritable relation de discours. Elle a été introduite dans ce modèle d'annotation pour traiter les segments correspondant à des adverbiaux détachés. Elle a pour argument l'adverbial détaché en tête de phrase ou en début de constituant (pour le segment cible), et le ou les segments simples qui sont indexés par cet adverbial, selon l'hypothèse d'encadrement de Charolles (1997). Ce second argument est souvent un segment complexe. Nous avons en (39) *Encadrement(a,[b-f])*. Comme nous l'avons présenté dans la sous-section 3.2.2, la SDRT traite les cadres de discours spatio-temporels comme des introducteurs de nouveaux topiques, reliés par la relation d'*Élaboration* au segment complexe composé des constituants sous la portée de l'introducteur de cadre. La possibilité de traiter les adverbiaux détachés au moyen de la relation d'*Élaboration* a été écartée car cela nécessite une bonne compréhension du modèle et la gestion des topiques semi-implicites. Par ailleurs, les types d'adverbiaux détachés segmentés sont plus larges que les seuls spatio-temporels pour lesquels il a été décrit leur rôle d'introducteur de nouveaux topiques.

- (39) [A la fin des années 1980,]_a [un consensus semble se dégager en France autour de l'idée de la fin de ce que l'on appelait l'«exception française».]_b [Les clivages idéologiques irréconciliables]_c [et l'atmosphère de guerre civile larvée ont disparu,]_d [tandis que les grands conflits sociaux,]_e [mais aussi l'engagement et la participation politique, sont en net déclin.]_f

Enfin, la **relation de localisation temporelle** n'est pas non plus une relation de discours. Elle relie deux segments si le segment à attacher permet seulement de localiser temporellement l'événement ou l'état décrit dans le segment cible comme en (40). Généralement le segment à attacher est une subordonnée temporelle ou un adverbial temporel comportant un nom d'éventualité.

- (40) [Paul est arrivé]_a [après le départ de Marie]_b

Relations d'élaboration

Deux relations de discours ont été retenues dans le modèle d'annotation : la relation d'*Élaboration* et la relation d'*Élaboration d'entité*. Nous citons le manuel pour ces deux relations centrales dans notre thèse.

★★★

La **relation d'Élaboration** relie deux propositions si la seconde proposition décrit un sous-état ou sous-événement de l'état ou de l'événement décrit dans la première proposition. La relation d'Elaboration inclut également les cas d'exemplification, de reformulation et de paraphrase.

Exemples• *Élaboration* entre événements

- (41) [Cette année-là vit de nombreux changements dans la vie de nos héros.]_a [Jean épousa Adèle,]_b [Marie s'acheta une maison à la campagne,]_c [et Paul partit pour le Brésil.]_d

Élaboration(*a*, [*b-d*])

- (42) [La Lausitz, une région pauvre de l'est de l'Allemagne,]_a [réputée pour ses mines de charbon à ciel ouvert,]_b a été le théâtre d'une première mondiale, mardi 9 septembre.]_c [Le groupe suédois Vattenfall a inauguré, dans la petite ville de Spremberg, une centrale électrique à charbon expérimentale qui met en IJuvre toute la chaîne des techniques de captage et de stockage du carbone (CCS).]_d

Élaboration d'entité (*c*, [*a-b*]) % *Lausitz*

Élaboration(*c*, *d*)

• *Élaboration* entre états

- (43) [Un soir,]_a [il faisait un temps horrible,]_b [les éclairs se croisaient,]_c [le tonnerre grondait,]_d [la pluie tombait à torrent ;]_e

Encadrement (*a*, *b*)

Élaboration (*b*, [*c-e*])

• Reformulation

- (44) [Comme son banquier si inspiré,]_a [le groupe Allard était en fâcheuse posture financière,]_b [c'est-à-dire au bord de la déconfiture.]_c

Encadrement (*a*, *b*)

Élaboration(*b*, *c*)

Marqueurs

Le plus souvent, *Élaboration* apparaît sans marqueur. Seuls les cas d'*Élaboration* qui sont des exemplifications ou des reformulations apparaissent parfois avec les marqueurs suivants : *par exemple, notamment, c'est à dire, à savoir*.

Confusions possibles

• Arrière-Plan

Arrière-Plan relie un état à un événement, et l'état ne fait pas partie de cet événement, il décrit un aspect de la scène dans laquelle l'événement se déroule, même si cette scène concerne des participants de l'événement (cf. Marie, portant un chapeau, est entrée dans le bar). À l'inverse, *Élaboration* relie soit deux événements, soit deux états, dont l'un fait partie de l'autre. La distinction état/événement suffit donc à séparer *Élaboration* d'*Arrière-Plan*.

• *Élaboration d'entité*

Élaboration introduit des précisions sur un événement ou un état, mais ces précisions sont vues en termes de sous-événements ou sous-états. Lorsque la précision porte sur un participant de l'événement ou l'état, il s'agit plutôt d'une *Élaboration d'entité*. Là encore,

la distinction état/événement peut parfois aider, puisque la description d'une entité est un état, alors que le premier argument peut décrire un événement.

La **relation d'Élaboration d'entité** lie deux segments dont le second (celui qui est attaché) précise une propriété d'une des entités impliquées dans le premier segment (la cible). Cette précision peut être importante (e.g., identificatoire) ou marginale. Il peut être utile d'annoter en commentaire l'entité concernée par l'élaboration.

Exemples

- (45) [Mikhaïl Saakachvili, [le jeune et bouillant président géorgien]_a, avait besoin d'action pour sauver son régime]_b.
Élaboration d'entité(b,a) % Mikhaïl Saakachvili
- (46) [Pourquoi a-t-on abattu Paul Mariani, [cinquante-cinq ans]_d, [attaché au cabinet de M. François Doubin,]_e [ministre délégué]_f, [chargé du commerce et de l'artisanat]_g ?]_h
Élaboration d'entité(h,[d-e]) % Paul Mariani
Élaboration d'entité(e,[f-g]) % François Doubin

Bien qu'*Élaboration d'entité* apparaisse souvent sous la forme d'une apposition, comme dans les exemples précédents, on trouve également des cas avec des propositions indépendantes :

- (47) [Je m'apprêtais à en faire autant pour le bas de mon pantalon]_c [(je suis extrêmement soigneux de mes effets),]_d [quand apparut mon ami Axelsen.]_e [Mon ami Axelsen est un jeune peintre norvégien,]_f [plein de talent et de sentimentalité.]_g
Arrière-plan(e,c)
Commentaire(c,d)
Élaboration d'entité(e,[f-g]) % Axelsen

Certaines reformulations correspondent à des élaborations d'entité

- (48) [Elle paraît contredire cette forme de " fin de l'histoire " à la française,]_a [c'est-à-dire cette reconnaissance quasi généralisée des principes de l'alternance démocratique,]_b
Élaboration d'entité (a,b)
- (49) [Aux législatives de 2002,]_a [le scrutin uninominal majoritaire à deux tours ne permettait que des mesures incitatives,]_b [à savoir des pénalités financières portant sur l'aide publique de l'Etat aux partis politiques.]_c
Encadrement (a,b)
Élaboration d'entité (b,c)

Marqueurs

La relation n'a pas de marqueurs explicites, sauf dans le cas de reformulation : *c'est-à-dire*, *à savoir*. En revanche, la relation a des indices : le segment attaché est souvent

une apposition (syntagme nominal ou adjectival), ou bien une relative explicative (non déterminative). Lorsqu'il s'agit d'une clause indépendante, le verbe est un verbe d'état dans une forme non progressive, et c'est un prédicat qui décrit une propriété stable, non transitoire.

Confusions possibles

•Cas d'une apposition

Si celle-ci peut être déplacée en fin de phrase, ce n'est pas une *Élaboration d'entité*, mais un *Arrière-plan* voire une *Explication* (ou les deux à la fois).

- (50) [Marie, [portant un chapeau,]_a entra dans le bar.]_b
Arrière-plan (b,a)
- (51) [Jean, [malade comme un chien,]_a est resté au lit aujourd'hui.]_b
Explication (b,a)
Arrière-plan (b,a)

Si l'apposition ne peut pas être déplacée vers la droite, et s'il s'agit d'un participe passé qui décrit l'état résultant d'un événement, et que cet événement est saillant (e.g., présence d'un modifieur temporel sur l'événement), alors il faut privilégier *Flashback*.

- (52) [Les pâtes de verre historiques du Musée de l'Ecole de Nancy, [volées en 1985]_c [et récupérées en juin 1990 au Japon par la police française]_d, vont-elles repartir dans ce pays?]_e
Flashback (e,[c-d])
Narration (c,d)

Dans les autres cas, on considérera qu'on a bien une *Élaboration d'entité*.

•Cas des relatives ou clauses indépendantes

Pour avoir une *Élaboration d'entité*, il faut que le segment attaché décrive un état et pas un événement. Si cet état est introduit par un verbe qui n'est pas à un temps progressif, et que l'état décrit une propriété stable de l'entité (être grand, blond, avoir la charge de...), alors on a une *Élaboration d'entité*.

- (53) [Alors qu'elle traversait le boulevard,]_a [la victime a été renversée par un homme de la brigade motocycliste de la direction de l'ordre public et de la circulation (DOPC)]_b [qui a notamment la charge des escortes officielles.]_c
Arrière-plan(b,a)
Élaboration d'entité (b,c) % brigade motocycliste de la DOCP

Dans les cas contraires, soit l'état ne décrit pas une propriété «individuelle» d'une entité, mais une propriété transitoire de la situation dans laquelle se trouve cette entité (X était assis dans son fauteuil, fatigué...), on est probablement en présence d'un arrière-Plan, voire d'une autre relation (causale, etc.).

Étant donné qu'un segment qui s'attache avec *Élaboration d'entité* introduit un état et qu'il peut s'attacher à un segment qui introduit un événement, il est possible de confondre *Élaboration d'entité* avec *Arrière-plan*. Beaucoup d'arrière-plans relient un segment à l'avant

plan introduisant un évènement et un autre segment avec un état comme l'exemple classique :

- (54) [Marie entra dans la cuisine.]_a [Paul faisait la vaisselle.]_b
Arrière-plan (*a,b*)

•Arrière-plan

Il faut insister sur le fait que la règle de déclenchement pour *Élaboration d'entité* est plus spécifique que celle pour *Arrière plan* - c'est-à-dire même si on a deux segments reliés où l'un introduit un évènement et l'autre un état, si le segment introduisant l'état décrit une propriété stable ou même définitoire d'un objet dans l'autre segment, il s'agit d'une *Élaboration d'entité* plutôt que d'un *arrière-plan*.

- (55) [La section judo de l' OFP a organisé ce week-end le premier tour de la compétition
 [appelée " les Petits Tigres "]_b , dans la salle du COSEC .]_c [Ce premier tour
 regroupait tous les petits judokas du secteur de Pompey nés'en 1994 - 95 - 96 et
 97]_d
Élaboration d'entité (*c,b*)
Élaboration d'entité(*c,d*)

Dans l'exemple ci-dessus on pourrait considérer *Arrière-plan*(*c,d*), sauf que la propriété introduite dans (d) sur «le premier tour» est définitoire. Donc nous avons *Élaboration d'entité*.

•Élaboration

La confusion avec la relation d'*Élaboration* ne devrait pas se produire lorsque la cible décrit un évènement, en effet, dans ce cas, puisque le segment attaché est un état, on ne peut avoir *Élaboration*(un état ne peut faire partie d'un évènement). Il convient cependant de noter que l'entité élaborée dans une *Élaboration d'entité* peut être une éventualité (par exemple un évènement) introduite dans le segment cible par un nominal. Cette éventualité n'est alors pas l'éventualité principale du segment-cible, il ne peut donc pas y avoir confusion.

- (56) [Il était 20 heures, samedi,]_a [quand les sapeurs-pompiers sont intervenus pour
 maîtriser un incendie]_b [qui avait pris dans une cave de la résidence Estoup, au 4,
 rue André-Breyer, à Tarbes (Hautes Pyrénées).]_c [Le sinistre était d'importance.]_d
Arrière-plan(*b,a*)
Élaboration d'entité (*b,[c-d]*) % incendie

Attention toutefois à ne pas confondre appositions et ellipses. Certaines reformulations sont en fait des ellipses, et dans ce cas, il s'agit de cas d'*Élaboration*, et non pas d'*Élaboration d'entité*. Toutefois, nous avons constaté des cas difficiles. Considérez l'attachement de (j) ou (k) à (h).

- (57) [Cette animation , [qui comporte trois tours ,]_h [dont le prochain aura lieu à Faulx
 [(date à définir)]_i]_j, [et le troisième le 1er mai, à Nomeny ,]_k est destinée à
 montrer une première approche de la compétition à tous ces petits sportifs .]_l
Élaboration d'entité (*h,j*)

Élaboration d'entité (h,k)
Élaboration d'entité (l,h)

L'attachement de (l) à (h) paraît une *Élaboration d'entité* standard. Par contre la relation entre (h) et (j) est beaucoup plus difficile à déterminer. Des verbes comme «avoir lieu» n'introduisent intuitivement pas d'événements séparés de ceux introduits par leurs sujets, même s'ils sont au futur. Celui-ci n'a pas de fonction autre que de préciser des localisations spatio-temporelles du sujet. Alors l'événement principal de (j) est l'événement dénoté par «le prochain». Celui-ci ne peut pas être un sous-événement de l'état de comporter. Alors il faudra annoter comme ci-dessus.

•Commentaire

La relation de *Commentaire* peut aussi décrire une propriété d'un objet mais cette propriété est toujours l'expression d'un jugement ou sentiment par rapport à cette entité de la part d'un agent. Ce ne sont donc pas des propriétés stables comme les cas prototypiques de l'*Élaboration d'entité*.

Relation de commentaire

Enfin, la **relation de Commentaire** est également une nouvelle relation par rapport à la SDRT, rajoutée pour les besoins de l'analyse. Un segment à attacher est relié à un segment cible si celui-ci donne le point de vue d'un agent ou de l'auteur sur ce qui est décrit dans le segment cible, comme en (58) où on a *Commentaire(a,[b,c])* et *But(b,c)*.

- (58) [... il convient de modifier le régime de priorité à cet endroit]_a. [La pose d'un panneau stop paraît être la formule la mieux adaptée]_b [pour assurer la sécurité des usagers]_c

Pour en terminer avec la présentation des relations utilisées dans le modèle d'annotation du projet ANNODIS, une dernière relation est employée, *Fusion*. Cette dernière n'est pas une relation de discours et a été introduite pour gérer des erreurs commises lors de la phase de segmentation. Cette relation permet de regrouper deux segments qui n'auraient pas dû être segmentés en premier lieu.

Nous souhaitons insister sur le fait que les résultats obtenus avec cette annotation ne sont pas un corpus annoté avec les relations de discours SDRT, bien qu'on perçoive l'influence du modèle dans les définitions des relations. Les annotateurs fournissent plutôt une **annotation intuitive** en s'appuyant sur les définitions données ci-dessus. On obtient des représentations des jugements des analystes *naïfs*.

8.2.2.3 Liste de marqueurs potentiels

Pour chaque relation, une liste de marqueurs potentiels est donnée à titre indicatif. Il est signalé que ces marqueurs sont ambigus, i.e. qu'ils peuvent signaler plusieurs relations de discours. Nous reviendrons sur cette liste lorsque nous aborderons les questions de l'exploitation des données annotées.

8.2.3 Outil d'annotation

La plateforme Glozz a spécialement été conçue dans le cadre du projet ANNODIS. Glozz est une plateforme générique d'aide à l'annotation manuelle de corpus.

Toutes les procédures d'annotation (perspective descendante et ascendante) ont été menées avec la plateforme Glozz. Toutefois, l'expérience d'annotation ascendante a révélé que cette plateforme n'était pas optimale pour une annotation en relations de discours. Nous travaillerons, ici, sur des annotations au format texte et sur des représentations graphiques de ces annotations (non fournies par la plateforme).

L'outil *Glozz* proposera également, au terme du projet, une interface d'interrogation.

8.2.4 Exploitation des données

L'interface d'interrogation de la plate-forme Glozz étant en cours de construction, nous avons, de notre côté, mis au point un programme Perl (appelé `generique-comparaison.pl`) nous facilitant une première exploitation des données issues du corpus ANNODIS.

Notre exploitation des données va essentiellement concerner le corpus de la partie ascendante. Pour rappel, nous avons 42 textes doublement annotés. Le programme `generique-comparaison.pl` permet de comparer, pour une relation donnée, un texte doublement annoté. Celui-ci permet de comparer toutes les annotations équivalentes.

Le programme extrait toutes les occurrences de cette relation dans les deux annotations concurrentes. Plusieurs configurations sont prises en compte :

- AA - AR (Accord Attachement - Accord Relation) Les deux annotateurs ont proposé une annotation identique ou comparable.
- AA - DR (Accord Attachement - Désaccord Relation) Les deux annotateurs ont attaché les mêmes segments mais ont proposé des relations différentes.
- DA (Désaccord Attachement) En cas de désaccord sur l'attachement des segments, aucune comparaison n'est possible concernant les relations de discours.

Pour compléter notre étude des structures énumératives, nous ferons appel aux annotations de la partie descendante et aux 18 extraits de textes également annotés par la perspective ascendante. Nous proposerons une comparaison manuelle des annotations proposées dans les deux approches (*Cf.* Chapitre 9, sous-section 9.2.2).

8.3 Perspectives

À partir d'une relation de discours, le programme fournit une liste de double annotations.

Il est ainsi possible de calculer les taux d'accord entre les deux annotateurs pour cette relation (fréquence AA-AR) et les relations les plus souvent en concurrence avec l'annotation de la relation concernée (AA-DR). Mais plus important encore, ce programme facilite l'accès à tous les segments reliés par la relation envisagée, permettant de porter un regard nouveau sur les réalisations linguistiques de celle-ci.

Dans le chapitre suivant, nous commençons par proposer une analyse quantitative des annotations de la relation d'*Élaboration*. Nous présentons la répartition de celle-ci dans le corpus, les taux d'accord et les relations qui lui sont le plus généralement concurrentes.

Puis, nous proposons une analyse qualitative des annotations, i.e. les annotations proposées par nos annotateurs correspondent-elles à des annotations d'*Élaboration* dans le standard ?

Enfin, pour une relation souvent signalée comme n'ayant pas de marqueurs prototypiques, nous proposons de porter un regard nouveau sur les réalisations linguistiques de celle-ci pour faire émerger de nouveaux indices de la relation et également valider les marqueurs que nous avons mis au jour.

Chapitre 9

Élaboration au cœur du corpus ANNODIS

Sommaire

9.1	Analyse quantitative	256
9.1.1	Relations les plus fréquentes	256
9.1.2	Accord entre annotateurs sur les relations d'élaboration	256
9.1.3	Divergences	258
9.2	Analyse qualitative	259
9.2.1	<i>Via</i> les marqueurs de la partie 2	259
9.2.1.1	Les adverbess paradigmatissants	260
9.2.1.2	Les marqueurs d'ouverture	269
9.2.1.3	Le gérondiff	272
9.2.1.4	Bilan	273
9.2.2	<i>Via</i> l'annotation des structures énumératives	275
9.2.2.1	Identité des structures selon les deux perspectives	275
9.2.2.2	Différence des structures entre les deux perspectives	279
9.2.2.3	Bilan	284
9.2.3	<i>Via</i> l'annotation standard	285
9.2.3.1	Analyse des accords entre annotateurs (AA-AR)	285
9.2.3.2	Analyse des divergences (AA-DR)	289
9.2.3.3	Analyse des désaccords (DA)	297
9.2.3.4	Rappel et précision	298
9.2.3.5	Bilan et perspectives	299
9.2.4	Bilan	299
9.3	Conclusion	300

Parallèlement aux travaux sur les marqueurs du discours que nous avons présentés dans la partie 2, nous avons participé, tout au long de nos quatre années de thèse, à la construction du corpus ANNODIS, pour la partie ascendante, annoté en relations de discours. Nous ne présentons pas une étude complète de l'exploitation du corpus, seulement disponible depuis début 2010, mais nous proposons, dans ce chapitre, une méthodologie

d'exploitation du corpus permettant d'en tirer parti pour nos recherches sur la relation d'*Élaboration*.

Nous proposons deux volets à cette étude : une étude quantitative et une étude qualitative.

L'étude quantitative (section 9.1) permet de situer la place de la relation d'*Élaboration* dans le corpus. Elle permet une première évaluation de la tâche d'annotation. Les taux d'accord (fréquence AA-AR), dont nous allons voir qu'ils sont plutôt médiocres, montrent la difficulté de cette tâche. Enfin, cette étude permet d'ouvrir de nouvelles pistes de recherche, à travers la mise au jour des relations les plus souvent en divergence avec la relation d'*Élaboration*.

Nous complétons cette étude avec une étude qualitative (section 9.2). Cette étude cerne plus directement les objectifs visés dans cette thèse, i.e. les réalisations linguistiques de la relation d'*Élaboration*. Nous abordons cette étude selon trois axes : (i) *via* les marqueurs du discours (sur les 42 textes doublement annotés) ; *via* la comparaison des annotations des approches ascendante et descendante sur les structures énumératives (sur les 18 textes concernés) et *via* l'annotation standard, i.e. notre évaluation¹ des paires concernées par une annotation d'*Élaboration* (AA-AR et AA-DR) (sur 16 textes).

9.1 Analyse quantitative

9.1.1 Relations les plus fréquentes

Comme nous le voyons dans le tableau 9.1, la relation d'*Élaboration* est la relation la plus fréquente dans les textes (30% des relations annotées). La relation d'*Élaboration d'entité* se trouve juste derrière avec 16% des occurrences. Une telle fréquence pour ces deux relations dans le corpus ANNODIS justifie l'intérêt qui leur est porté dans cette thèse.

9.1.2 Accord entre annotateurs sur les relations d'élaboration

L'intérêt porté à la relation d'*Élaboration* se justifie non seulement par la fréquence de cette relation mais également par les taux d'accord entre annotateurs, dont nous allons voir avec les tableaux 9.2 et 9.3 qu'ils sont plutôt médiocres.

Les deux tableaux 9.2 et 9.3 se lisent de la façon suivante : pour les textes doublement annotés par «annot1» et «annot2», la seconde colonne donne la fréquence à laquelle la relation a été annotée par «annot1» et la troisième colonne donne la fréquence à laquelle la relation a été annotée par «annot2». La dernière colonne donne le nombre d'annotations communes de la relation (AA-AR). Par exemple, dans le tableau 9.2, dans les textes annotés par Sophie et Nadej, Sophie a annoté 193 fois la relation d'*Élaboration* et Nadej 224 fois. 90 annotations sont communes chez nos deux annotateurs.

¹Nous nous sommes basée sur l'annotation standard de 11 textes proposée par les membres de la partie ascendante du projet ANNODIS. Chaque annotation standard fait l'objet d'une concertation entre deux membres du projet. Pour éviter la confusion entre les références aux exemples et aux segments, les références aux exemples se feront toujours entre parenthèses () et les références aux segments se feront entre crochets <> dans ce chapitre.

Relation de discours	Annot1 : Nadej	Annot2 : Simon	Annot3 : Sophie	Total
Élaboration	480	309	411	1200
Élaboration d'entité	82	360	183	625
Continuation	87	72	232	391
Encadrement	110	77	45	232
Contraste	70	83	43	196
Narration	57	90	33	180
Explication	70	44	43	157
Résultat	61	53	39	153
But	35	64	35	134
Commentaire	32	24	72	128
Attribution	19	35	36	90
Arrière-Plan	20	30	23	73
Parallèle	9	31	10	50
Localisation temporelle	17	2	12	31
Flashback	6	8	16	30
Alternation	9	7	8	24
Conditionnel	6	13	5	24
Totaux	1170	1302	1246	3718

TAB. 9.1 – Fréquences des relations de discours dans les 42 textes du corpus ANNODIS

Couple d'annotateurs Annot1/Annot2	Nombre d'annotations d'Élaboration Annot1	Nombre d'annotations d'Élaboration Annot2	Nombre d'annotations d'Élaboration communes (AA-AR)
Sophie/Nadej	193	224	90
Nadej/Simon	197	146	85
Simon/Sophie	130	192	59
Totaux	1082		234

TAB. 9.2 – Nombre d'annotations d'Élaboration par annotateur et nombre d'annotations communes de celle-ci

Couple d'annotateurs Annot1/Annot2	Nombre d'annotations d'Élaboration d'entité Annot1	Nombre d'annotations d'Élaboration d'entité Annot2	Nombre d'annotations d'Élaboration d'entité communes (AA-AR)
Sophie/Nadej	38	78	20
Nadej/Simon	43	147	37
Simon/Sophie	186	96	71
Totaux	588		128

TAB. 9.3 – Nombre d'annotations d'Élaboration d'entité par annotateur et nombre d'annotations communes de celle-ci

Nous évaluons l'accord entre annotateurs en nous basant sur la fréquence d'annotations communes de la relation concernée. Nous notons que cette fréquence est basse pour les relations d'*Élaboration* et d'*Élaboration d'entité*. D'une part, on note une différence dans le nombre de relations annotées d'un annotateur à l'autre (par exemple, dans le tableau 9.3, Simon annoté plus de 3 fois plus d'*Élaboration d'entité* que Nadej). D'autre part, on note que, dans la majorité des cas, moins de la moitié des annotations d'un annotateur correspond à une annotation commune.

Cet accord faible entre annotateurs est révélateur, selon notre point de vue, de la difficulté de la tâche d'annotation et plus précisément, pour ce qui est de notre objet d'étude, la difficulté d'annoter la relation d'*Élaboration*. Nous pensons que cet accord faible justifie l'intérêt porté à la relation. Il s'agit de tenter de déterminer les difficultés rencontrées par les annotateurs (incompréhension générale de la tâche, incompréhension des descriptions des relations de discours...) et de pallier ces difficultés pour de futures campagnes d'annotation. Ce travail peut notamment se faire en étudiant, dans les cas d'accord sur l'attachement mais de désaccord sur la relation de discours (AA-DR), l'annotation divergente.

9.1.3 Divergences

Dans un second volet de notre programme d'exploitation des données, nous allons chercher à découvrir, en cas d'accord sur l'attachement mais de désaccord sur la relation, quelle est la relation divergente (AA-DR). Dans le tableau 9.4, voici la liste des relations de discours qui sont le plus souvent en divergence avec l'annotation de la relation d'*Élaboration* et de la relation d'*Élaboration d'entité*.

Dans la majorité des divergences, *Élaboration* est en concurrence avec une autre relation subordonnante.

Sans surprise, nous voyons que dans presque 40% des cas, la relation d'*Élaboration* est en divergence avec la relation d'*Élaboration d'entité* et dans 70% des cas, la relation d'*Élaboration d'entité* est en divergence avec la relation d'*Élaboration*.

Concernant plus particulièrement la relation d'*Élaboration*, celle-ci est en divergence, dans presque 10% des cas, avec la relation d'*Encadrement*. Ce résultat n'est pas totalement une surprise. D'un côté, il est vrai qu'il ne paraît pas compliqué de reconnaître un adverbial détaché et de l'attacher par la relation d'*Encadrement* à son contexte droit. De l'autre côté, ces annotations sont en parfait accord avec les travaux sur les adverbiaux spatio-temporels détachés proposés par (Vieu *et al.*, 2005).

La divergence avec la relation d'*Explication* est observée dans 8 % des cas.

La divergence d'annotation avec *Arrière-plan* a déjà été mise en évidence à plusieurs reprises et a même mené Prévot *et al.* (2009) à introduire la nouvelle relation d'*Élaboration d'entité*. Cette divergence demeure dans 4% des cas.

La divergence avec la relation de *Continuation*, dans presque 10% des cas, est plus surprenante d'un point de vue théorique, *Élaboration* étant subordonnante et *Continuation* coordonnante.

Nous reviendrons sur chacune de ces divergences dans la sous-section 9.2.3.2.

Annotation concurrente d'Élaboration	Nombre de cas	Annotation concurrente d'Élaboration d'entité	Nombre de cas
Élaboration d'entité	161	Élaboration	151
Encadrement	41	Continuation	15
Continuation	39	Commentaire	6
Explication	36	Contraste	5
Commentaire	25	Arrière plan	5
Résultat	20	Explication	6
Arrière-Plan	18	Localisation temporelle	4
Contraste	16	Résultat	4
Attribution	13	Narration	2
But	12	Encadrement	2
Narration	7	Attribution	1
Localisation temporelle	7		
Parallèle	5		
Flashback	4		
Alternation	1		
Conditionnel	1		

TAB. 9.4 – Annotations divergentes avec les relations d'Élaboration et d'Élaboration d'entité

9.2 Analyse qualitative

9.2.1 Via les marqueurs de la partie 2

Nous proposons d'abord notre analyse qualitative par le biais des marqueurs décrits dans la partie 2.

Nous souhaitons, d'une part, valider ou invalider nos analyses à la lumière de ces nouvelles données. Après avoir adopté une démarche inductive dans la partie 2, partant de l'observation de nos marqueurs pour établir des hypothèses sur leurs fonctionnements discursifs et les valider, nous nous situons à présent dans une démarche déductive, visant à valider nos observations sur les marqueurs de l'Élaboration avec de nouvelles données.

Nous nous intéressons, d'autre part, au rôle que ces marqueurs ont joué dans la tâche d'annotation. Cela nous mène à porter une réflexion sur le rôle des marqueurs et indices dans une perspective d'annotation. Nous rappelons qu'une liste de marqueurs potentiels a été fournie dans le manuel. Nous listons dans le tableau 9.5 ceux qui ont été mis en corrélation avec la relation d'Élaboration.

Pour chaque marqueur présent dans un segment, nous nous référons aux annotations proposées par nos annotateurs de ce segment à son contexte gauche. Nous allons focaliser notre attention sur les trois points suivants en nous demandant :

1. si le segment-cible de la relation annotée est identique chez les deux annotateurs (AA) ;
2. si la relation de discours annotée est identique chez les deux annotateurs (AR) et

Marqueurs	Relations possibles
c'est-à-dire	Élaboration
de même	Parallèle, Continuation, Élaboration
de plus	Continuation, Élaboration
d'ailleurs	Commentaire, Élaboration
en effet	Élaboration, Explication
en particulier	Élaboration
en plus	Continuation, Élaboration
en plus de	Continuation, Élaboration
en résumé	Élaboration (inversée)
notamment	Élaboration
par exemple	Élaboration

TAB. 9.5 – Liste de marqueurs potentiels pour la relation d'Élaboration

3. si la relation de discours du standard valide les analyses que nous avons fournies.

9.2.1.1 Les adverbess paradigmatissants

Dans les 42 textes doublement annotés, nous avons relevé 16 occurrences de *notamment*, 2 de *en particulier*, 3 de *plus précisément* et 4 de *plus particulièrement*. Concernant *notamment* et *en particulier*, les résultats des annotations doivent être évidemment nuancé par le fait qu'ils ont été signalés aux annotateurs comme marqueurs potentiels de la relation d'Élaboration, Cf. Tableau 9.5.

•Notamment

Dans 10 cas sur 16, le segment contenant *notamment* est attaché au contexte gauche par Élaboration par les deux annotateurs.

L'exemple (1) fait partie de ces 10 cas :

- (1) [La France a également [(juillet 2007)]₄₄ publié une Stratégie nationale d'adaptation au changement climatique]₄₃ [et envisagerait une gouvernance adaptée,]₄₅ [**notamment** dans le cadre du Grenelle de l'Environnement.]₄₆ (rechauffementClimatique_7_2.seg)
AA-AR : Élaboration (45,46)

Élaboration relie le segment <46> au segment <45>. Dans les 10 cas, les annotateurs s'accordent sur la relation et sur le segment-cible. Ce résultat positif doit être considéré à sa juste valeur, d'abord parce que *notamment* a été indiqué comme un marqueur de la relation d'Élaboration (ce qui a certainement influencé les annotateurs) et deuxièmement parce que l'attachement de deux segments consécutifs appartenant à une même phrase (dans 9 cas) ou à la phrase précédente (dans 1 cas de *notamment* en position initiale) ne pose pas vraiment de problème.

L'annotation standard distingue deux cas. Pour les premiers cas (5 occurrences), l'annotation standard est en accord avec l'annotation d'Élaboration comme en (1) et (2).

- (2) [Ces créances douteuses, [lorsqu'elles deviennent importantes,]₆₁ sont une véritable gangrène de l'économie]₆₀ [- **notamment** dans les pays en voie de développement]₆₂ [où l'euphorie de la croissance a tendance à faire oublier aux banques les règles élémentaires du contrôle des risques.]₆₃ (wk417-pib)
 AA-AR : *Élaboration* (60,62)

En (2), le segment <62> est syntaxiquement un modifieur de la proposition en <60>. Il est apposé, séparé par un tiret et introduit par *notamment* et fait ainsi l'objet d'une segmentation. Du point de vue de l'interprétation et suivant la description de la relation donnée dans le manuel, on peut interpréter <62> comme contenant une ellipse et «être une véritable gangrène de l'économie dans les pays en voie de développement» est un sous-état d'«être une véritable gangrène». Le segment <62> apporte des indications spatiales à l'interprétation du segment <60>.

Pour les seconds cas (5 cas), l'annotation standard propose *Élaboration d'entité*, en lien avec la description de celle-ci dans le manuel comme en (3) et (4).

- (3) [Plusieurs milliers de personnes blessées et des milliers d'autres, [**notamment** parmi les secouristes,]₃₆ sont atteintes de maladies engendrées par l'inhalation de poussières toxiques.]₃₅ (attentats.seg)
 AA-AR : *Élaboration* (35,36)
- (4) [Ces attentats ont généré des effets puissants et persistants,]₂₉ [**notamment** politiques et économiques.]₃₀ (attentats.seg)
 AA-AR : *Élaboration* (29,30)

Dans ces deux exemples, le segment introduit par *notamment* est syntaxiquement un modifieur d'un élément de la phrase : en (3), il modifie le nom «autres» et en (4), il modifie le nom «effets». Suivant la définition de la relation d'*Élaboration d'entité*, les segments <36> en (3) et <30> en (4) précisent une ou des propriétés d'une des entités impliquées dans le segment-cible. L'annotation standard propose donc *Élaboration d'entité* (35,36) pour (3) et *Élaboration d'entité* (29,30) pour (4).

Dans le dernier cas d'AA-AR, la relation de *But* est proposée. On peut penser que les annotateurs naïfs se sont appuyés, pour cette annotation, sur la présence du marqueur *pour*.

- (5) [La Directive sur le système européen d'échange de droits d'émission sera modifiée en 2008,]₁₄ [pour inclure **notamment** les émissions de l'aviation.]₁₅ (rechauffement-Climatique_7_2.seg)
 AA-AR : *But* (14,15)

Dans les 5 cas d'AA-DR, un cas concerne la divergence entre *Élaboration* et *Explication*, illustré en (6).

- (6) [La récente Histoire de la littérature belge parue chez Fayard [(Paris, 2003)]₈ estime]₇ [que ce Manifeste s'oppose trait pour trait à l'idée de belgitude]₉ [telle qu'elle avait été manifestée]₁₀ [notamment dans le numéro spécial de la revue de l'ULB en 1980 intitulé La Belgique malgré tout.]₁₁ [**Notamment** parce que les signataires du Manifeste faisaient entendre un "Nous" collectif]₁₂ [là où les collaborateurs de la revue

de l'ULB faisaient entendre autant de "je" individuels.]₁₃ (WK329-Reactions.seg)

Annot1 : *Élaboration*(9,12)

Annot2 : *Explication* ([7-11],[12-13])

En (6), nous avons deux annotations différentes. D'un côté le segment complexe <7-11> est attaché par *Explication* aux segments <12-13>. L'annotateur a suivi les instructions données par *parce que*. De l'autre côté, l'annotateur a proposé *Élaboration*(9,12). Il semblerait que celui-ci se soit laissé influencer par la liste des marqueurs. Pourtant l'inférence établie par *parce que* est, ici, prioritaire. Comme nous l'avons montré dans nos analyses, *notamment* indique qu'il y a d'autres explications, écartées par l'emploi de l'adverbe.

Les 4 autres cas de divergence concernent les relations d'*Élaboration* et d'*Élaboration d'entité*. Cette divergence n'est pas étonnante du fait que nous avons montré ci-dessus que *notamment* peut introduire un segment modifiant une éventualité principale (impliquant *Élaboration*) ou une entité (impliquant *Élaboration d'entité*).

L'exemple (7) est un cas où l'annotation standard propose *Élaboration* (comme dans 2 cas).

- (7) [C'est dans le Nouveau Testament que l'on trouve trace de Capharnaüm pour la première fois.]₁₃ [Cette ville est abondamment citée dans les récits de l'Évangile]₁₄ [où elle est désignée comme le lieu]₁₅ [où vécut Jésus de Nazareth durant une grande partie de sa vie publique en Galilée.]₁₆ [C'est **notamment** là qu'il aurait guéri de nombreux malades]₁₇ [et qu'il exorcisa plusieurs personnes possédées par des démons.]₁₈ (passage_wk89_capharnaum.seg)

Annot1 : *Élaboration*([13-16],[17-18])

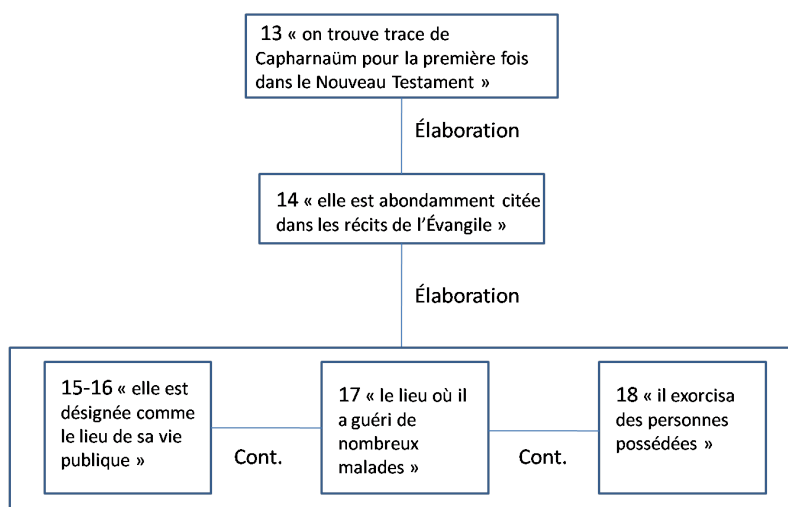
Annot2 : *Élaboration d'entité* (13,17)

En (7), le segment complexe <17-18> contenant *notamment* est attaché, selon les annotateurs, soit par *Élaboration* au segment complexe <13-16>, soit par *Élaboration d'entité* au segment <13>. *Notamment* introduit le déictique *là*, faisant référence anaphoriquement à la ville «Capharnaüm». On peut donc comprendre le segment <17> comme étant une élaboration de l'entité «Capharnaüm» du segment <13>. En revanche, l'attachement du segment <17> au segment <13> par *Élaboration* nous paraît plus difficilement interprétable.

Deux interprétations semblent possibles. La première est proposée par l'annotation standard. Les segments <17> et <18> sont des continuations de <15-16> : «Capharnaüm est connue pour avoir été le lieu de vie publique de Jésus mais aussi l'endroit où il a guéri et exorcisé des personnes», comme illustré en Figure 9.1.

Dans l'interprétation que nous proposons, «guérir des malades» et «exorciser des personnes possédées» sont considérés comme des sous-événements de la vie publique de Jésus <15-16>. On attache donc <15-16> à <14> par *Élaboration*, comme dans la première interprétation proposée. En revanche, on attache <17> et <18> par *Élaboration* à <15-16> comme illustré en Figure 9.2.

L'exemple (8) est un emploi initial de *notamment*. Nous prédisons donc une relation d'*Élaboration* entre <23> et <24-27>. Nos annotateurs ont proposé deux annotations concurrentes.

FIG. 9.1 – 1^{ère} interprétation de (7)

- (8) [Les Khazars surent bâtir une civilisation évoluée sur les plans technique et politique.]₂₃
 [Notamment,]₂₄ [ils frappaient monnaie]₂₅ [et possédaient la technologie du papier,]₂₆
 [héritée de leurs voisins chinois.]₂₇ (wk273-Khazar.seg)
 Annot1 : *Élaboration*(23,[24-27])
 Annot2 : *Élaboration d'entité* (23,[24-27])

Les éventualités contenues dans le segment complexe «frapper monnaie» et «posséder la technologie du papier» ne sont pas des sous-éventualités de «savoir bâtir une civilisation évoluée sur les plans technique et politique». Mais elles élaborent le résultat de ce dernier, à savoir «avoir une civilisation évoluée» obtenue par leur savoir. On peut ainsi comprendre les deux annotations, soit <24-27> élabore le résultat de «savoir bâtir», soit il élabore l'entité «civilisation évoluée».

De notre point de vue, les états introduits par *notamment* «ils frappaient monnaie» et «possédaient la technologie du papier» élabore plutôt l'état «ils avaient une civilisation évoluée» plutôt que l'entité «civilisation évoluée», allant dans le sens de la relation d'*Élaboration*.

En (9), nous avons la deuxième occurrence de *notamment*, notée ici *notamment*₃₈, qui se trouve dans un segment dont l'annotation est divergente.

- (9) [Son œuvre, [en tant qu'écrivain,]₃₄ comporte des récits et des écrits sur l'art]₃₃
 [(notamment sur le Cubisme et l'Art informel)]₃₅ [mais c'est surtout [pour ses
 essais sur le langage et la littérature]₃₇ qu'il est célèbre,]₃₆ [notamment Les Fleurs
 de Tarbes ou la Terreur dans les lettres, À demain la poésie, Petite Préface à toute
 critique.]₃₈ (passage_wk257_jeanpaulhan.seg)
 Annot1 : *Élaboration* (37,38)
 Annot2 : *Élaboration d'entité* (37,38)

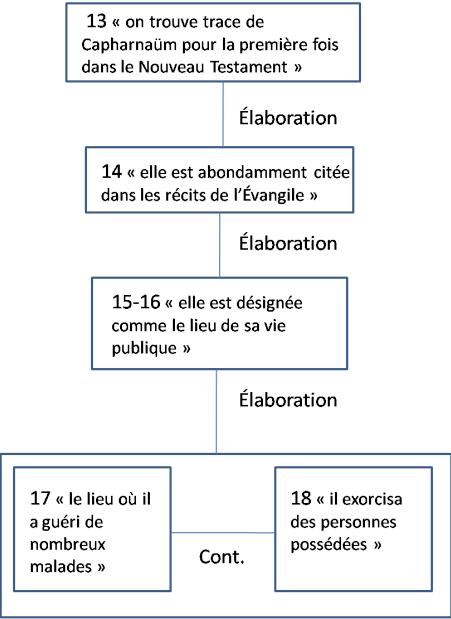


FIG. 9.2 – 2^{ième} interprétation de (7)

L’annotation standard propose *Élaboration d’entité*. Le segment <38> introduit des précisions sur l’entité «essais» du segment <37>.

Nous récapitulons, dans le tableau 9.6, la répartition des annotations des naïfs et nos annotations pour *notamment*.

- En particulier
Tout comme *notamment*, *en particulier* a été signalé dans la liste du manuel comme marqueur potentiel de la relation d’*Élaboration*. Parmi les deux occurrences annotées de

Annotation des naïfs		Annotation standard	
AA-AR			
Élaboration	10	Élaboration	5
		Élaboration d’entité	5
But	1	But	1
AA-DR			
Élaboration <i>vs.</i> Explication	1	Explication	1
Élaboration <i>vs.</i> Élaboration d’entité	4	Élaboration	2
		Élaboration d’entité	2

TAB. 9.6 – Récapitulatif des annotations des naïfs et de l’annotation standard pour *notamment*

en particulier dans le corpus, nous avons un emploi interphrastique en (10) et un emploi à l'initiale en (11). Les annotateurs naïfs se sont accordés dans les deux cas avec l'annotation de la relation *Élaboration*, comme en (10).

- (10) [L'Inhibition de contact est le fait]₂ [que de nombreuses cellules normales [ou dérivant de cellules normales [(**en particulier** des cellules épithéliales),]₅ mises en culture,]₄ se multiplient]₃ [jusqu'à ce qu'elles aient formé une couche simple.]₆
(WK241-InhibitionDeContact.seg)
AA-AR : *Élaboration* (4,5)

En (10), les deux annotateurs ont attaché <5> à <4> par *Élaboration* alors que le segment <5> vient précisément donner des propriétés sur une entité impliquée dans le segment <4>, à savoir «cellules normales». La relation d'*Élaboration d'entité* est proposée dans l'annotation standard. Une fois de plus, nous pensons que les annotateurs ont été influencés par la liste de marqueurs, *en particulier* étant signalé comme un marqueur potentiel de la relation d'*Élaboration*.

En (11), nous avons un emploi à l'initiale de *en particulier*. Il introduit, comme attendu dans nos analyses, un segment élaborant.

- (11) [Les modèles les plus récents simulent de façon satisfaisante les variations de température au cours du xxe siècle.]₁₂ [**En particulier**,]₁₃ [les simulations menées sur le climat du xxe siècle sans intégrer l'influence humaine ne rendent pas compte du réchauffement climatique,]₁₄ [tandis que celles incluant cette influence sont en accord avec les observations.]₁₅ (Rechauffement_climatique_4_1.seg)
AA-AR : *Élaboration* (12,[13-15])

Les annotateurs naïfs attachent le segment complexe <13-15> au segment <12> par *Élaboration*. Le segment complexe <13-15> oppose les «modèles anciens» décrits en <13-14> (fusionnés) qui ne tiennent pas compte de l'influence humaine et les «modèles récents» décrits en <15> qui le font. <13-14> est introduit par *en particulier* qui opère une sorte de décrochement vis-à-vis de <12>, en signalant que des informations plus particulières, plus en détail, vont être fournies. Sans la locution adverbiale il est intéressant d'observer le changement d'interprétation. L'attachement de <14> à <12> se fait alors par la relation de *Contraste*, comme on peut le constater en (12)

- (12) [Les modèles les plus récents simulent de façon satisfaisante les variations de température au cours du xxe siècle.]₁₂ [Les simulations menées sur le climat du xxe siècle sans intégrer l'influence humaine ne rendent pas compte du réchauffement climatique,]₁₄ [tandis que celles incluant cette influence sont en accord avec les observations.]₁₅

Ensuite, le segment <15> serait logiquement rattaché à <14> par la relation de *Contraste* en s'appuyant sur la conjonction *tandis que*. On serait alors tenté de vouloir attacher <15> à <13> par *Élaboration* mais ce dernier n'est plus sur la frontière droite.

L'analyse de (11) renforce nos analyses présentées dans le chapitre 5. *En particulier* joue un rôle très important en bloquant l'inférence de la relation de *Contraste* et en imposant

Annotation des naïfs		Annotation standard	
AA-AR			
Élaboration	2	Élaboration	1
		Élaboration d'entité	1

TAB. 9.7 – Récapitulatif des annotations des naïfs et de l'annotation standard pour *en particulier*

l'inférence de la relation d'*Élaboration*. Ce rôle est primordial car ce n'est qu'en arrivant au segment <15> qu'on comprend le rôle de celui-ci.

Nous récapitulons, dans le tableau 9.7, la répartition des annotations des naïfs et les propositions faites par l'annotation standard pour *en particulier*.

• *Plus précisément*

Nous avons 3 occurrences de *plus précisément*. En (13), nous avons un cas de divergence d'annotation :

- (13) [Le départ de Mariana est maintenant connu;]₃₉ [la voie continue ensuite en direction du sud.]₄₀ [Elle passe par plusieurs communes]₄₁ [et, franchit le Golu, le Fium'altu, la rivière d'Alesani, la Bravone jusqu'au Tavignanu.]₄₂ [La distance proposée par l'Itinéraire entre Mariana et Aleria correspond à peu près à la distance actuelle.]₄₃ [Seule différence,]₄₄ [la route actuelle ne passe pas par Mariana,]₄₅ [mais à 6 Km à l'ouest de la cité.]₄₆ **[Plus précisément**, de Mariana au Fium'altu on peut considérer une voie relativement droite,]₄₇ [ce que confirme l'étude de la carte I.G.N de Vescovatu au 1/25000ème.]₄₈ [Cette voie passe par le lieu dit «Migliarine»]₄₉ [sur la commune de Castellare-di-Casinca;]₅₀ [ce toponyme signifie, [nous l'avons vu précédemment,]₅₂ une halte ou un relais.]₅₁ (wk521-mariana.seg)
Annot1 : Élaboration d'entité([44-46],47)
Annot2 : Encadrement (46,47)

D'un côté, un annotateur a relié <47> à <46> par la relation d'*Encadrement*, ce qui ne fait pas vraiment sens au vu de la définition de cette relation dans le manuel. De l'autre côté, un annotateur a relié <47> à un segment complexe formé des segments <44> à <46> par la relation d'*Élaboration d'entité*.

De notre point de vue, le segment <47> introduit par *plus précisément* introduit une redescription des segments «45-46». <45-46> et <47> décrivent de deux façons différentes un même tracé. *Plus précisément* indique que le second segment est une redescription du premier. Nous proposons donc *Élaboration*.

En (14) et (15), les annotateurs naïfs s'accordent sur l'annotation d'*Élaboration*.

- (14) [La musique n'adoucit pas toujours les mœurs.]₂ [Exemple à l'Harmonie municipale de Montbéliard,]₃ [qui a vu soudainement ses rangs s'éclaircir.]₄ **[Plus précisément**,]₅ [c'est la batterie [-une des deux composantes de l'Harmonie-]₇ qui connaît des problèmes.]₆ (er061-batterie.seg)
AA-AR : Élaboration([3-4],[5-6])

Annotation des naïfs		Annotation standard	
AA-AR			
Élaboration	2	Élaboration	2
AA-DR			
Élaboration d'entité <i>vs.</i> Encadrement	1	Élaboration	1

TAB. 9.8 – Récapitulatif des annotations des naïfs et de l'annotation standard pour *plus précisément*

- (15) [Une octave sur un clavier correspond à la distance entre deux notes successives du même nom,]₃₃ [entre do et do par exemple.]₃₄ [L'octave est **plus précisément** une différence de 6 tons entre deux notes.]₃₅ (passage_wk410_pianoclavier.seg)
AA-AR : *Élaboration* (33,35)

Ces exemples de *plus précisément*, deux en position initiale et l'autre en position intraphrastique, vont dans le sens des analyses fournies dans le chapitre 5. Comme nous l'avons décrit, *plus précisément* introduit une éventualité qui est interprétée comme une redescription de l'éventualité introduite dans le segment précédent. Ces données renforcent notre conclusion de considérer *plus précisément* comme un marqueur d'*Élaboration*.

Malheureusement, nous n'avons pas suffisamment de données pour apprécier le rôle que ce marqueur a pu jouer lors de l'interprétation dans la tête de nos annotateurs.

Nous récapitulons, dans le tableau 9.8, la répartition des annotations des naïfs et les propositions faites par l'annotation standard pour *plus précisément*.

• *Plus particulièrement*

Les 4 occurrences relevées de *plus particulièrement* sont des emplois en position initiale. Le segment introduit par *plus particulièrement* est dans 3 cas sur 4 attaché par *Élaboration* au contexte gauche. En revanche, le point d'attachement diffère sensiblement entre les deux annotateurs. Deux stratégies se font face : d'un côté, un annotateur propose de rattacher le segment introduit par *plus particulièrement* à un segment complexe, de l'autre côté, un annotateur propose de rattacher le segment introduit par *plus particulièrement* à un segment (contenu à l'intérieur du segment complexe proposé par le premier annotateur). Nous pouvons généralement trancher entre l'une ou l'autre des deux annotations proposées.

En (16), le segment introduit par *plus particulièrement* doit être attaché à <12> et non au segment complexe formé de <11-12> reliés par *Attribution*. Les segments fusionnés <14-15> n'élaborent pas l'événement «révéler» en <11>. En revanche l'information fournie en <14-15>, «l'inhibition de contact n'a aucun effet sur les cellules cancéreuses» est une information plus précise que celle fournie en <12> «les signaux qui font normalement cesser la croissance n'ont aucun effet sur les cellules cancéreuses», en s'appuyant sur la connaissance de l'inhibition de contact comme un signal faisant normalement cesser la croissance des cellules.

- (16) [Une étude des cellules cancéreuses en culture a révélé]₁₁ [que les signaux qui font normalement cesser la croissance n'ont aucun effet sur elles.]₁₂ [Plus particulièrement,]₁₄

[les cellules cancéreuses en culture sont insensibles à l'inhibition de contact.]₁₅
 (WK241-InhibitionDeContact.seg)
Annot1 : *Élaboration*([11-12],[14-15])
Annot2 : *Élaboration*(12,[14-15])

Pour (17), les segments introduits par *plus particulièrement*, i.e. tous des items illustrant les faiblesses du PIB, sont attachés soit au segment <11>, soit au segment complexe formé des segments de <10> à <13>. La deuxième annotation nous paraît la plus convenable étant donné que les items dominés par *plus particulièrement* sont des illustrations soit du cloisonnement décrit en <12>, soit de la limite à la sphère du compte de résultat décrit en <11> et <13>.

- (17) [Les critiques mettent ainsi en avant deux faiblesses du PIB :]₁₀ [son cloisonnement comptable [(ce qui échappe à la comptabilité nationale échappe au PIB)]₁₂ et sa limite à la sphère du compte de résultat]₁₁ [qui ignore l'actif et le passif.]₁₃ [Plus particulièrement :]₁₄
 [- Le PIB ne tient pas compte de l'auto-consommation,]₁₅ (...)
 [- Il ne tient pas compte du travail au noir]₂₀ [(par définition non comptabilisé)]₂₁
 [dont la contribution à l'économie peut être très importante dans certains pays.]₂₂
 [- Il ne tient pas compte de la valeur estimée des actifs et passifs [(le patrimoine)]₂₄
 publics et privés,]₂₃ [ni des externalités positives ou négatives]₂₅ [qui font évoluer cette valeur]₂₆ [et donc contribuent à un gain ou à une perte de moyens.]₂₇ (...)
 (wk417-pib.seg)
Annot1 : *Élaboration*(11,14)
Annot 2 : *Élaboration*([10-13],14)

Enfin pour l'exemple (18), les mêmes segments sont mis en relation par les deux annotateurs, le segment complexe formé des segments de <2> à <7> et le segment complexe formé des segments de <8> à <13>. Mais un annotateur a relié ces deux segments complexes par *Élaboration* et l'autre par *Élaboration d'entité*.

- (18) [On appelle panchromatique [(du grec pan, tout et chrôma, couleur)]₃ une réponse physico-chimique [(émulsion photographique, pigment photosensible, vision, etc.)]₄ qui ne discrimine pas les couleurs,]₂ [c'est-à-dire dont le processus est identique]₆ [quelle que soit la longueur d'onde de la lumière incidente.]₇ [Plus particulièrement,]₈ [un support sensible est dit panchromatique]₉ [lorsqu'il est sensible à toutes les longueurs d'onde du spectre visible,]₁₀ [à contrario d'un support orthochromatique,]₁₁ [lequel n'est sensible qu'au bleu et au vert,]₁₂ [tandis qu'un film non-chromatisé sera sensible seulement au bleu.]₁₃ (wk385-panchromatique.seg)
Annot1 : *Élaboration*([2-7],[8-13])
Annot2 : *Élaboration d'entité* ([2-7],[8-13])

Comme décrit dans nos analyses, ici *plus particulièrement* met en relation une information générale «on appelle panchromatique une réponse physico-chimique qui ne discrimine pas les couleurs» et une information particulière «un support sensible est dit panchromatique lorsqu'il est sensible à toutes les couleurs». Le premier segment donne une définition générale au terme de «panchromatique» et le second, introduit par *plus particulièrement* donne une information plus précise en donnant une définition de «support sensible panchromatique». Nous avons proposé, dans le chapitre 5, de traiter ces cas au moyen de la relation d'*Élaboration*.

Annotation des naïfs		Annotation standard	
AA-AR			
Élaboration	3	Élaboration	3
AA-DR			
Élaboration <i>vs.</i> Élaboration d'entité	1	Élaboration	1

TAB. 9.9 – Récapitulatif des annotations des naïfs et de l'annotation standard pour *plus particulièrement*

Dans les 4 occurrences de *plus particulièrement* observées, l'annotation standard propose, dans tous les cas, la relation d'*Élaboration*. Aucun cas ne présente la seconde configuration décrite dans le chapitre 5, pour laquelle *plus particulièrement* est attaché par une relation coordonnante au contexte gauche. Ces données nous confirment sensiblement que *plus particulièrement* est un indice intéressant pour la recherche de la relation d'*Élaboration* en corpus.

Nous récapitulons, dans le tableau 9.9, la répartition des annotations des naïfs et notre annotation pour *plus particulièrement*.

9.2.1.2 Les marqueurs d'ouverture

Encore moins de données sont disponibles pour les marqueurs de l'organisation du discours : 2 occurrences de *d'abord* et 2 occurrences de *dans un premier temps*.

• *D'abord*

Dans l'exemple (19), les annotateurs ont proposé la même annotation. Le segment <43> contenant *d'abord* interne a été attaché au segment <42> par la relation d'*Élaboration*. Puis le segment <44> par *Narration* au segment <43>. Les segments <46> et <47> ont été annotés comme des continuations de <44>.

- (19) [Elle a beaucoup voyagé]₄₁ [et surtout changé d'employeur.]₄₂ [Chez Peugeot **d'abord**,]₄₃
 [puis successivement à Héricourt]₄₄ [(filature),]₄₅ [ouvreuse de cinéma,]₄₆ [vendeuse
 de magasins]₄₇ [(maroquinerie à Montbéliard),]₄₈ [avant de revenir au contrôle chez
 Peugeot.]₄₉ (extr_estrep253_MaisonBélot.seg)
 AA-AR : *Élaboration* (42,43)

Nous proposons la même structure du discours, illustré Figure 9.3.

Les fonctionnements en termes de relation de discours de *d'abord* interne n'ont pas, à notre connaissance, été étudiés. Bien que non nécessaire pour l'interprétation de la relation d'*Élaboration*, en (19), son rôle est tout de même de faciliter celle-ci.

L'exemple (20) est un cas de divergence sur la relation. Les deux relations sont subordonnantes (*Élaboration* et *Explication*) allant dans le sens des analyses proposées pour *d'abord*, marqueur de relation subordonnante.

- (20) [Habitée très tôt par la passion du chant,]₆ [elle fait partie de plusieurs troupes,]₇
 [**d'abord** celle de Fadela Dziria]₈ [où elle assure les chœurs en jouant du tambourin,]₉

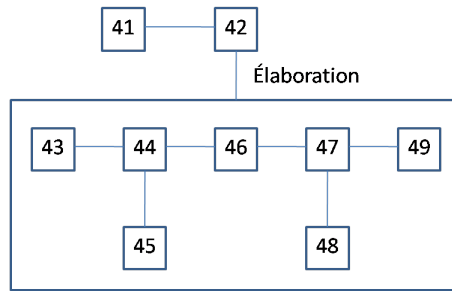


FIG. 9.3 – Structure du discours de (19)

Annotation des naïfs		Annotation standard	
AA-AR			
Élaboration	1	Élaboration	1
AA-DR			
Élaboration <i>vs.</i> Explication	1	Élaboration	1

TAB. 9.10 – Récapitulatif des annotations des naïfs et de l’annotation standard pour *d’abord*

[puis une autre]₁₀ [qu’elle dirige avec sa complice Flifla,]₁₁ [enfin la sienne]₁₂ [où elle est la chanteuse principale]₁₃ [et devient une animatrice réputée des fêtes de mariages.]₁₄

Annot1 : *Élaboration* (7,[8,10,12])

Annot2 : *Explication* (7,8)

La relation d’*Explication* ne semble pas du tout appropriée. Une divergence possible serait, semble-t-il, entre *Élaboration* et *Élaboration d’entité*. En (20), faut-il considérer les segments <8-9>, <10-11> et <12-13> comme venant spécifier l’entité «troupes» ou comme venant spécifier l’état «faire partie de plusieurs troupes» ?

À première vue, élaborer l’entité «troupes» ou élaborer «faire partie de plusieurs troupes» permet d’accéder à deux représentations du discours semblables. En revanche, si on interprète les segments <8-13> comme décrivant les troupes dont Biyouna a fait partie de façon successive, i.e. si les segments <8-9>, <10-11> et <12-13> sont reliés par la relation de *Narration*, seule la relation d’*Élaboration* est possible.

Nous récapitulons, dans le tableau 9.10, la répartition des annotations des naïfs et de l’annotation standard pour *d’abord*.

• *Dans un premier temps*

Dans le cas de nos 2 occurrences de *dans un premier temps*, les annotations présentent des divergences au niveau de l’attachement.

Bras et Schnedecker (2009) ont décrit le rôle de *dans un premier temps* comme introduisant la première étape d’une élaboration. Mais dans certains cas, un topique de discours explicite est élaboré par le segment contenant *dans un premier temps* et dans les autres,

un topique de discours implicite est construit, puis élaboré par le segment contenant *dans un premier temps*.

Dans l'exemple (21), un topique de discours semi-explicite est disponible en <5>.

- (21) [Il y a trois semaines,]₅ [l'Uruguayen envisageait de quitter le club.]₆ [**Dans un premier temps**,]₇ [le président nancéien invita Pablo à cesser sa carrière]₈ [pour intégrer le staff technique.]₉ (extr_estrep025_PabloCorrea.seg)
Annot1 : Résultat(5,[7-8])
Annot2 : (? , 7)

L'interprétation de cet exemple fait intervenir des inférences complexes. Le segment <5> joue le rôle d'un cadre de discours, qui dans le cadre de la SDRT, déclenche l'introduction d'un constituant semi-explicite contenant un événement sous-spécifié du type «ce qui s'est passé il y a trois semaines». Le segment <6> introduit un arrière-plan à cet événement sous-spécifié (Vieu *et al.*, 2010). Enfin, les segments <7-9> décrivent le «premier temps» de cet événement sous-spécifié. On les attache donc à <5> par *Élaboration*.

Nos annotateurs naïfs, quant à eux, ne disposent pas des outils de la SDRT et des notions de constituants semi-implicites ou implicites. De ce fait, l'attachement des segments introduits par *dans un premier temps* n'est pas facilité pour nos annotateurs. Un annotateur propose un rattachement au segment <5> par *Résultat*. Mais un résultat sur un segment introduisant une localisation temporelle ne fait pas vraiment sens. Le second annotateur ne trouve, quant à lui, pas de segment dans le contexte gauche lui convenant pour l'attachement des segments introduits par *dans un premier temps*.

Le même problème se pose pour l'exemple suivant.

- (22) [Lors du dépassement d'un véhicule circulant dans le sens Chavannes-sur-l'Etang - Montreux-Vieux,]₇ [le côté gauche de la voiture [qui doublait]₉ a mordu l'accotement.]₈ [L'automobile a perdu sa roue gauche,]₁₀ [et a décollé]₁₁ [pour se retourner]₁₂ [et terminer sa course sur le toit.]₁₃ [Le conducteur, [Benattia Grabis,]₁₅ [24 ans,]₁₆ [de Cernay,]₁₇ [ainsi que ses deux occupants,]₁₈ [Abdel Rhazza,]₁₉ [22 ans,]₁₄ [de Retzwiller]₂₀ [et Lionel Ledoucin,]₂₁ [30 ans,]₂₂ [de Foussemagne,]₂₃ [ont été gravement blessés dans l'accident.]₂₄ [Dans un premier temps,]₂₅ [les occupants de la voiture doublée ont porté secours aux victimes.]₂₆ [L'état de ces dernières a fait dépêcher sur place des ambulances des pompiers de Montreux-Château, d'Alkirch et de Dannemarie]₂₇ [ainsi que les Samu du Haut-Rhin et du Territoire de Belfort.]₂₈ [Les blessés ont été évacués sur le centre hospitalier de Belfort.]₂₉ (extr_estrep049_Tonneaux.seg)
Annot1 : Continuation ([7-10],[25-26])
Annot2 : Élaboration(24,25)

Dans un premier temps introduit la première étape d'une réaction faisant suite à l'accident. Selon l'analyse de Bras et Schnedecker (2009), son attachement nécessite l'introduction d'un topique implicite du type «ce qui s'est passé après l'accident», résultant de l'accident décrit dans les segments <7-24>. N'ayant pas la possibilité d'introduire ce segment implicite, le premier annotateur propose un rattachement par *Continuation* aux segments décrivant l'accident <7-10> et le second annotateur un attachement au segment décrivant les blessés de l'accident <24> par *Élaboration*.

Annotation des naïfs		Annotation standard	
AA-DR			
Résultat <i>vs.</i> Pas d'attachement	1	Élaboration	2
Continuation <i>vs.</i> Élaboration	1		

TAB. 9.11 – Récapitulatif des annotations des naïfs et de notre annotation pour *dans un premier temps*

Nous pouvons penser que les divergences d'annotation sont révélatrices des inférences complexes à l'œuvre en présence des cadres de discours entraînant l'introduction d'un constituant semi-implicites (comme illustré en (21) avec le segment <5>) et du marqueur *dans un premier temps*, entraînant, en (22), l'introduction d'un topique implicite, inférences complexes mises au jour par des analystes experts.

Nous récapitulons, dans le tableau 9.11, la répartition des annotations des naïfs et notre annotation pour *dans un premier temps*.

9.2.1.3 Le gérondif

Nous relevons 3 des configurations de constructions gérondives que nous avons présentées dans le chapitre 7. Seuls 9 gérondifs ont été segmentés alors que le corpus en contient 18. Malheureusement pour nous, les annotateurs, lors de la phase de segmentation, n'ont pas segmenté les 9 autres gérondifs (*Cf.* les spécifications du manuel de segmentation dans le chapitre 8 dans lesquelles il est précisé que tous les gérondifs doivent être segmentés).

Nous allons nous focaliser sur les constructions gérondives que nous analysons au moyen de la relation d'*Élaboration* (ou de *Circonstance accompagnante* que nous avons décrit comme un sous-type d'*Élaboration*) (5 cas).

Dans 3 cas, les annotateurs naïfs ont également proposé la relation d'*Élaboration*, comme illustré en (23).

- (23) [Chomsky a commencé à développer sa théorie de la grammaire générative et transformationnelle dans les années 1950]₉ [**en cherchant** à dépasser aussi bien l'approche structuraliste, distributionnaliste que béhavioriste dans l'étude du langage naturel.]₁₀ (chomsky.seg)
AA-AR : Élaboration (9,10)

Dans 2 cas (24) et (25), nous avons une annotation divergente.

- (24) [Le gouvernement américain a institué une législation sécuritaire]₃₁ [et, [**en dénonçant** un nouvel «Axe du Mal»,]₃₃ s'est lancé dans une «guerre contre le terrorisme»].₃₂ (attentats.seg)
Annot1 : Élaboration(32,33)
Annot 2 : Arrière-plan (32,33)

Le segment <33> décrit une manière «dénoncer un nouvel Axe du Mal» d'accomplir l'action décrite en <32> «se lancer dans une guerre contre le terrorisme». Nous proposons donc la relation d'*Élaboration*.

Annotation des naïfs		Notre annotation	
AA-AR			
Élaboration	3	Élaboration	3
AA-DR			
Élaboration <i>vs.</i> Arrière-plan	1	Élaboration	1
Élaboration <i>vs.</i> Continuation	1	Circonstance Accompagnante	1

TAB. 9.12 – Récapitulatif des annotations des naïfs et de notre annotation pour le gérondif

- (25) [Nommé Maestro,]₆₃ [il permet d’ouvrir la bouteille avec facilité,]₆₄ [tout en conservant le bruit caractéristique des anciens bouchons.]₆₅ (bouchon.seg)
Annot1 : Continuation (64,65)
Annot2 : Élaboration(64,65)

Sans surprise, au vu de nos analyses, un annotateur a proposé la relation coordonnante *Continuation* et l’autre la relation subordonnante *Élaboration* pour relier les segments <64> et <65>. Nous avons décrit la complexité de ces cas de gérondifs dans le chapitre 7, pour lesquels nous avons proposé d’introduire une nouvelle relation de discours, *Circonstance accompagnante*. Les deux annotations divergentes semblent correspondre chacune à un aspect de la relation de discours à l’œuvre dans cet exemple. L’indice *tout en + gérondif* est pertinent pour filtrer les constructions gérondives traitées au moyen de *Circonstance accompagnante* de celles traitées au moyen d’*Élaboration*.

Nous récapitulons, dans le tableau 9.12, la répartition des annotations des naïfs et de notre annotation pour le gérondif.

9.2.1.4 Bilan

Premièrement, l’examen de ces exemples et de leurs annotations ne remet pas en question les analyses de la partie 2. Deuxièmement, il permet de tirer des conclusions sur le rôle de nos marqueurs dans l’identification de la relation d’*Élaboration*. Les occurrences de nos marqueurs ici observées dépassent le cadre de nos analyses. Nous avons regardé les occurrences de *notamment* et de *d’abord* dans toutes les positions de la phrase tandis que les analyses présentées dans la partie 2 n’ont eu pour objet que certains de leurs emplois.

En restant vigilant, au vu du nombre d’occurrences de ces derniers, nous considérons les adverbes paradigmatiques considérés ainsi que les marqueurs structurels considérés et la construction gérondive comme des indices pertinents pour l’identification de la relation d’*Élaboration*, tel qu’on peut le voir dans notre tableau récapitulatif 9.13.

En théorie, mettre en corrélation l’emploi d’un indice (un item lexical ou une construction syntaxique) avec l’annotation en relations de discours devrait suffire pour désigner ce dernier comme un indice fiable ou ambigu d’une relation de discours particulière : si l’indice est toujours en corrélation avec une seule relation de discours, il est considéré comme un indice fiable (marqueur) de celle-ci, si l’indice est en corrélation avec plusieurs relations de

Annotation des naïfs		Notre annotation	
Notamment (16 occurrences)			
AA-AR Élaboration	11	Élaboration	5
		Élaboration d'entité	6
AA-DR Élaboration	5	Élaboration	2
		Élaboration d'entité	2
		Explication	1
En particulier (2 occurrences)			
AA-AR Élaboration	2	Élaboration	2
Plus précisément (3 occurrences)			
AA-AR Élaboration	2	Élaboration	2
AA-DR Autres RD	1	Élaboration	1
Plus particulièrement (4 occurrences)			
AA-AR Élaboration	3	Élaboration	3
AA-DR Élaboration	1	Élaboration	1
D'abord (2 occurrences)			
AA-AR Élaboration	1	Élaboration	1
AA-DR Élaboration	1	Élaboration	1
Dans un premier temps (2 occurrences)			
AA-DR Autres RD	2	Élaboration	2
Le gérondif (9 occurrences)			
AA-AR Élaboration	3	Élaboration	3
AA-DR Élaboration	2	Élaboration	1
		Circonstance Accompagnante	1

TAB. 9.13 – Récapitulatif

discours, il est considéré comme un indice ambigu (on peut également établir les fréquences à laquelle il apparaît avec chacune des relations de discours).

En pratique, pour le corpus ANNODIS, la quantité des données annotées est insuffisante pour établir ce type de corrélation. Par ailleurs, la qualité de ces données est également insuffisante, ce qu'a montré la confrontation des annotations aux analyses sémasiologiques de notre partie 2. Ces données présentent des insuffisances tant au niveau de la segmentation que de l'annotation en relations de discours.

De plus, les analyses de la partie 2 ainsi que les observations faites sur les données du corpus ont montré que ce n'est généralement pas un indice mais plutôt un faisceau d'indices qui doit être mis en corrélation avec une ou plusieurs relations de discours. Des annotations en relations de discours ne peuvent pas mettre au jour ces faisceaux, seule une analyse sémasiologique semble pouvoir le permettre.

9.2.2 *Via l'annotation des structures énumératives*

La double perspective d'annotation du corpus ANNODIS nous permet d'amorcer une analyse de la rencontre des perspectives descendante et ascendante à partir des structures énumératives. Pour cette analyse, nous employons les 18 textes qui ont été annotés à la fois par les annotateurs de la perspective ascendante et ceux de la perspective descendante.

Dans les annotations de la perspective descendante, nous avons relevé 16 structures énumératives. Notre objectif est d'observer la façon dont ces structures ont été annotées par les annotateurs de l'approche ascendante et si ces dernières sont isomorphiques à des structures élaboratives.

Dans 12 structures énumératives, l'annotation de l'approche descendante peut être rapprochée de celle de l'approche ascendante. Dans les 4 autres, on relève des différences sur les structures annotées dans les deux perspectives.

9.2.2.1 *Identité des structures selon les deux perspectives*

• Structure énumérative simple

Les marques typo-dispositionnelles des structures énumératives rendent saillantes et facilement identifiables les structures énumératives. En présence de ces marques, les annotateurs de la perspective descendante et ascendante ont proposé des structures qui peuvent être rapprochées, comme illustré avec l'exemple (26) et les Figures 9.4 et 9.5.

- (26) [On distingue nettement, [sur un bouchon,]₃₀ de par la différence d'aspect des lièges :]₂₉
 [- La tête, [parfois appelée manche et partiellement engagée dans le goulot,]₃₂ est constituée d'un aggloméré de liège.]₃₁ [Cette méthode de fabrication autorise des tailles de bouchons bien supérieures]₃₃ [à ce qu'il est possible de récolter aujourd'hui sur les arbres.]₃₄ [De moindre qualité]₃₅ [(pour le vin),]₃₆ [cette partie n'est pas en contact avec le vin.]₃₇
 [- Le corps, [constitué de deux rondelles de liège massif aux plis croisés,]₃₉ est collé à la base de la tête.]₃₈ [En contact avec le vin,]₄₀ [cette partie assure l'étanchéité du flacon.]₄₁

[On distingue nettement, sur un bouchon, de par la différence d'aspect des lièges.] ₂₉	amorce
♦ [La tête, parfois appelée manche et partiellement engagée dans le goulot, est constituée d'un aggloméré de liège.] ₃₁ [Cette méthode de fabrication autorise des tailles de bouchons bien supérieures à ce qu'il est possible de récolter aujourd'hui sur les arbres.] ₃₄ [De moindre qualité (pour le vin), cette partie n'est pas en contact avec le vin.] ₃₇	item
♦ [Le corps, constitué de deux rondelles de liège massif aux plis croisés, est collé à la base de la tête.] ₃₈ [En contact avec le vin, cette partie assure l'étanchéité du flacon.] ₄₁	item

FIG. 9.4 – Annotation descendante de (26)

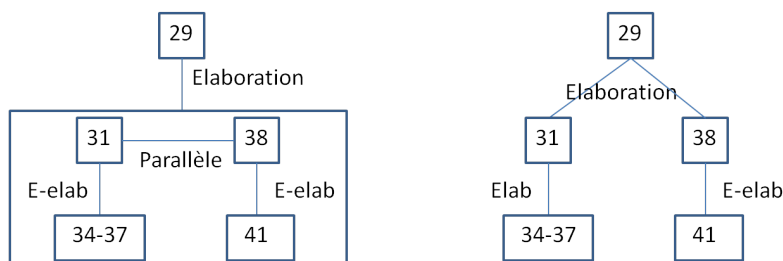


FIG. 9.5 – Annotations ascendantes quasi équivalentes de (26)

Cet exemple illustre les deux méthodes d'analyse de catégorisation (de la perspective descendante) et de mise en relation (de la perspective ascendante). Les segments, catégorisés comme des items, forment, dans la perspective ascendante, un segment complexe relié au segment catégorisé comme amorce par la relation d'*Élaboration*. Les items élaborent, selon les deux annotateurs naïfs, la description du bouchon annoncée dans l'amorce.

Cela indique que ces derniers ont, en quelque sorte, reconstruit un contenu pour le segment <29>.

Ce type d'amorce est dite *incomplète*. Cette incomplétude se manifeste au niveau de la syntaxe, i.e. le complément du verbe «distinguer» n'est pas donné dans la phrase amorce.

Si l'on suit une analyse SDRT et suivant notre proposition du chapitre 6, la relation d'*Énumération* peut être inférée grâce à la reconnaissance de l'amorce et des marques typodispositionnelles indiquant les items. Seule la relation d'*Énumération* peut être inférée sur la base de ces indices. Pour inférer *Élaboration*, il faut, en revanche, que le contenu des items remonte dans le constituant introduit pour l'amorce du type : «on distingue, sur un bouchon, deux parties». Les annotations de nos annotateurs naïfs semblent, par ailleurs, indiquer qu'ils ont, inconsciemment, reconstruit ce type de topique. Une fois la relation d'*Énumération* inférée, le topique peut être mis à jour et la relation de contenu *Élaboration* inférée dans un second temps. L'analyse SDRT est schématisée Figure 9.6.

• Structure énumérative à deux temps

3 structures énumératives, que nous identifions comme des structures énumératives à deux temps, ont été annotées dans les 18 textes. Dans la structure énumérative à deux temps en (27), deux principes d'organisation sont interdépendants : la première énuméra-

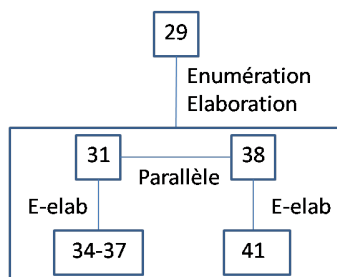


FIG. 9.6 – Représentation SDRT de (26)

tion est introduite par des marqueurs typo-dispositionnels et le marqueur lexical *ou bien* et la seconde énumération est introduite par des corrélats anaphoriques *dans le premier cas* et *dans le second cas* qui renvoient anaphoriquement aux items de la première énumération.

- (27) [Pour reprendre l'exemple tant débattu du Cours de linguistique générale et de son rôle dans l'histoire des idées linguistiques contemporaines, il me semble qu'on a aujourd'hui deux manières de considérer son statut :]₁₃
- [ou bien on considère que c'est le Cours qui a effectivement joué un rôle séminal dans la genèse des différents structuralismes comme si le texte possédait en lui-même et de manière virtuelle son historicité, le principe de son devenir ;]₁₄₋₁₆
 - [ou bien on cherche la productivité historique de ce texte dans la manière dont on y a renvoyé, dont on s'y est référé en cherchant à caractériser le plus précisément possible les «modes de références» et les reconstructions dont il a été l'objet dans des contextes scientifiques, culturels, institutionnels les plus divers.]₁₇₋₂₀
- [Dans le premier cas, on voit bien que la référence à un texte du passé ne nous fait pas quitter un présentisme profondément anhistorique : tout tient dans la lecture d'un texte et dans les lectures de lectures qui en ont été faites, la question de la littéralité du texte se confondant avec celle de sa vérité et rien n'empêchant de penser qu'une bonne lecture ou un retour au vrai texte (celui qu'on nous avait caché, qu'on avait défiguré, qu'on a retrouvé par hasard, qu'on va retrouver...), permettra de faire retour au Saussure authentique, et permettra également de dessiner pour la linguistique une avenir dont elle avait été abusivement privée. Il ne suffit plus alors qu'à s'auto-proclamer le héraut de cet avenir indéfiniment ouvert... (cf. F. Rastier - 2004 , P. Bouissac - 2001 qui l'un et l'autre plaident pourtant pour un retour «historique» à Saussure. Avec quel sens ici tu terme histoire ?).]₂₁₋₃₃
- [Dans le second cas, il s'agira plutôt de considérer le texte de 1916 comme une matrice projective et productive, assez puissante pour ordonner une série de projets scientifiques apparentés (la linguistique des Cercles, le structuralisme «généralisé», les sémiologies, etc.) apparentés, mais aussi assez fondamentalement concurrents et dispersés. (Pour ce point de vue «minoritaire», cf. Puech 2000 et 2005, Trabant, 2005).]₃₄₋₄₁

La structure proposée par l'approche descendante (Figure 9.7) ne rend pas compte de

cette organisation. La perspective descendante peut mettre en évidence l'amorce et les quatre items mais n'a pas prévu de séparer les deux temps de l'énumération, qui ne se situent pourtant pas au même niveau, comme peut le laisser entendre leur annotation.

[Pour reprendre l'exemple tant débattu du <i>Cours de linguistique générale</i> et de son rôle dans l'histoire des idées linguistiques contemporaines, il me semble qu'on a aujourd'hui deux manières de considérer son statut :]. ₁₃	amorce
- [ou bien on considère que c'est le <i>Cours</i> qui a effectivement joué un rôle séminal dans la genèse des différents structuralismes comme si le texte possédait <i>en lui-même</i> et de <i>manière virtuelle</i> son historicité, le principe de son devenir ;]. ₁₄₋₁₆	item
- [ou bien on cherche la <i>productivité historique</i> de ce texte dans la manière dont on y a renvoyé, dont on s'y est référé en cherchant à caractériser le plus précisément possible les « modes de références » et les reconstructions dont il a été l'objet dans des contextes scientifiques, culturels, institutionnels les plus divers.]. ₁₇₋₂₀	item
[Dans le premier cas, on voit bien que la référence à un texte du passé ne nous fait pas quitter un <i>présentisme</i> profondément anhistorique : tout tient dans la lecture d'un texte et dans les lectures de lectures qui en ont été faites, la question de la <i>littéralité</i> du texte se confondant avec celle de sa <i>vérité</i> et rien n'empêchant de penser qu'une <i>bonne</i> lecture ou un retour au <i>vrai</i> texte (celui qu'on nous avait caché, qu'on avait défiguré, qu'on a retrouvé par hasard, qu'on va retrouver...), permettra de faire retour au Saussure <i>authentique</i> , et permettra également de dessiner pour la linguistique une avenir dont elle avait été abusivement privée. Il ne suffit plus alors qu'à s'auto-proclamer le héraut de cet avenir indéfiniment ouvert... (cf. F. Rastier - 2004 , P. Bouissac – 2001 qui l'un et l'autre plaident pourtant pour un retour « historique » à Saussure. Avec quel sens ici tu terme histoire ?).]. ₂₁₋₃₃	item
[Dans le second cas, il s'agira plutôt de considérer le texte de 1916 comme une <i>matrice projective et productive</i> , assez puissante pour ordonner une série de projets scientifiques apparentés (la linguistique des Cercles, le structuralisme « généralisé », les sémiologies, etc.) apparentés, mais aussi assez fondamentalement concurrents et dispersés. (Pour ce point de vue « minoritaire », cf. Puech 2000 et 2005, Trabant, 2005).]. ₃₄₋₄₁	item

FIG. 9.7 – Annotation descendante de (27)

Comme nous l'avons montré dans nos analyses des structures énumératives à deux temps, les corrélats anaphoriques sont des informations fortes, présentant les items de la première énumération comme disponibles pour la poursuite du discours, ce que nos annotateurs ascendants ont perçu les amenant tout naturellement à traverser la frontière droite pour le rattachement du premier item de la seconde énumération au premier item de la première, comme illustré en Figure 9.8. Nos annotateurs ascendants ont également annoté la relation d'*Élaboration* à l'œuvre entre l'amorce et les items.

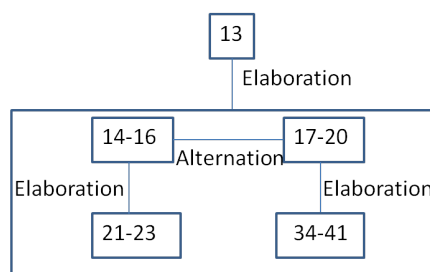


FIG. 9.8 – Annotation ascendante des annotateurs naïfs de (27)

Suivant une analyse de type SDRT, l'élaboration est, ici, annoncée dans l'amorce : «il y a deux manières de considérer le statut du *Cours de linguistique générale*». La première énumération est signalée par des marqueurs typo-dispositionnels. La seconde est signalée par des corrélats anaphoriques, qui renvoient aux items de la première énumération. Suivant nos propositions faites dans le chapitre 6, le second temps est directement attaché à l'amorce par *Énumération* et par *Élaboration*, par *Énumération* afin que les corrélats anaphoriques puissent faire référence sans ambiguïté aux items de la première énumération selon l'ordre dans lequel ils ont été introduits («dans le premier cas» signifiant «dans le premier cas que j'ai exposé») et par *Élaboration* pour rendre compte du fait que cette seconde énumération poursuit l'élaboration des manières de considérer le statut du *Cours de linguistique générale*. Nous rappelons ici que cet attachement du second temps par *Énumération* à l'amorce évite également de transgresser le principe de la frontière droite.

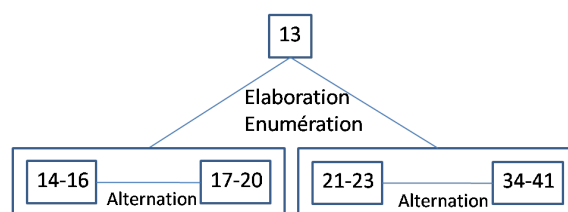


FIG. 9.9 – Représentation SDRT de (27)

9.2.2.2 Différence des structures entre les deux perspectives

Lorsque la structure énumérative n'est pas explicitement signalée par des marqueurs typo-dispositionnels (les deux points, les tirets...) comme en (28), les annotations proposées dans la perspective descendante ne suivent pas la structure proposée par les annotateurs de la perspective ascendante, comme illustré en Figures 9.10 et 9.11 :

- (28) [C'est en Afrique que les plus vieux ossements ont été découverts]₃₅. [Aujourd'hui,]₃₆
 [les paléontologues donnent à Homo sapiens un âge d'environ 200 000 ans]₃₇
 [puisque les plus vieux ossements retrouvés sont deux crânes datés de -195 000
 ans,]₃₈ [et appelés Omo 1 et Omo 2]₃₉ ; [viennent ensuite ceux de l'homme d'Herto
 encore appelé Homo sapiens idaltu,]₄₀ [datés d'environ -154 000 ans.]₄₁ [Ensuite
 viennent les ossements de Qafzeh et Skhul en Israël/Palestine datés respectivement
 de -97 000 et -80 000 ans.]₄₂
 [Les plus célèbres sont ceux de l'homme de Cro-Magnon,]₄₃ [datés de -35 000 ans
 et découverts en France.]₄₄

Comme nous pouvons le constater, les deux structures ne sont pas isomorphes. Du côté de l'annotation descendante, on a le segment <35> qui est considéré comme phrase introductrice de l'énumération <36-44>. Du côté de l'annotation ascendante, le bloc <36-44> n'a pas été attaché au reste de la structure. L'adverbial détaché *aujourd'hui* introduit dans le segment <36> domine les segments <37-44>, ce qui nous semble l'analyse appropriée. Le segment <37> indique que le plus vieil Homo sapiens découvert a un âge d'environ

amorce	[C'est en Afrique que les plus vieux ossements ont été découverts] ₃₅	[Aujourd'hui,] ₃₆	item
	[les paléontologues donnent à Homo sapiens un âge d'environ 200 000 ans] ₃₇		
	[puisqu'il y a les plus vieux ossements retrouvés sont deux crânes datés de -195 000 ans,] ₃₈	[et appelés Omo 1 et Omo 2] ₃₉ ;	item
	[viennent ensuite ceux de l'homme d'Herto encore appelé Homo sapiens idaltu,] ₄₀	[datés d'environ -154 000 ans.] ₄₁	
	[Ensuite viennent les ossements de Qafzeh et Skhul en Israël/Palestine datés respectivement de -97 000 et -80 000 ans.] ₄₂		item
	[Les plus célèbres sont ceux de l'homme de Cro-Magnon,] ₄₃	[datés de -35 000 ans et découverts en France.] ₄₄	item

FIG. 9.10 – Annotation descendante de (28)

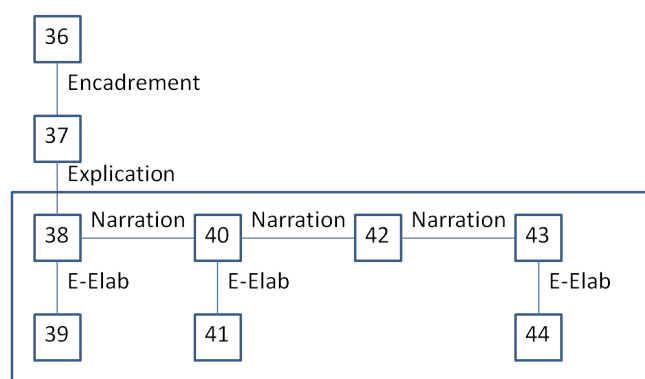


FIG. 9.11 – Annotation ascendante d'un annotateur naïf de (28)

200000 ans, le segment <38> apporte une preuve de cette assertion. Enfin, dans l'annotation ascendante naïve, les segments <40-41>, <42> et <43-44> remontent le temps en indiquant les datations des autres ossements correspondant à la période de l'Homo sapiens. Mais ces derniers ne sont plus des preuves de <37>. De ce fait, nous proposons une autre annotation, Cf. 9.12. Les segments <40-41> poursuivent la description de la découverte des ossements en Afrique, ils sont ainsi attachés à <37>. Le segment <42> décrit les ossements découverts en Israël/Palestine. Il est donc attaché à <35>. Enfin, les segments <43-44> décrivent les ossements découverts en France. Ils sont donc attachés à <42>. Nous sommes d'avis de ne pas considérer ici les segments <36-39>, <40-41>, <42> et <43-44> comme faisant partie d'une structure énumérative, du moins de celle proposée par l'annotation descendante. On peut penser que l'annotateur descendant, qui parcourt le texte à la recherche de structures énumératives, s'est laissé influencé par les deux occurrences du marqueur lexical *ensuite* et la structure en paragraphes des deux derniers items.

Nous proposons la structure donnée en Figure 9.12. Dans celle-ci, nous voyons que l'on peut considérer une énumération sans amorce correspondant à la mention de trois pays : «en Afrique», «en Israël/Palestine» et «en France» constituant des topiques reliés par *Continuation*.

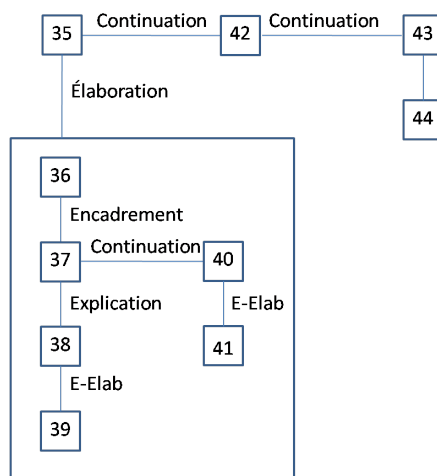


FIG. 9.12 – Notre annotation pour (28)

Les deux structures proposées par les analyses descendante et ascendante de l'exemple (29), en Figures 9.13 et 9.14, mettent en évidence deux organisations différentes du texte.

- (29) [Vinification]1 [Le vin de Champagne est produit selon la méthode traditionnelle, jadis appelée méthode champenoise, qui consiste principalement à opérer une double fermentation du moût, la première en cuves, la seconde dans les bouteilles mêmes, en cave, avec remuage régulier.(...)]2–14
 [La première fermentation, appelée fermentation alcoolique est identique à celle que subissent les vins tranquilles (c'est-à-dire non effervescents). Elle peut être suivie, mais ce n'est pas toujours le cas, d'une fermentation malolactique.(...)]15–30
 [En début d'année (qui suit la récolte), les vins sont suffisamment clairs pour être goûtés et procéder à l'étape de l'assemblage qui mélange en proportions variant à chaque année des vins de cépages, terroirs et millésimes différents (aucune autre A.O.C. en France ne permet ce type de mélange de vin de différents millésimes).]31–38
 [Quoi qu'il en soit, au moment d'embouteiller le vin de base ainsi obtenu, on lui ajoute la liqueur de tirage, composée de levures et de sucre. Cela enclenchera la dernière fermentation, dite prise de mousse.]39–44 C'est cette deuxième fermentation qui va donner naissance aux bulles de dioxyde de carbone. La bouteille est alors bouchée avec une capsule métallique analogue à celle des bouteilles de bière.]45–46
 [Cependant, cette deuxième fermentation produit des lies abondantes dont on devra débarrasser le vin par la suite.]47–49
 [Suit alors la période de vieillissement du vin en bouteilles d'une année environ pour les non millésimés à trois ans et plus pour les bouteilles millésimées.]50
 [Après ce vieillissement, on rangeait autrefois les bouteilles sur des étagères appelées «pupitres» où elles étaient penchées le goulot vers le bas. Chaque jour, les bouteilles étaient remuées, c'est-à-dire tournées d'un quart de tour, d'un mouve-

ment sec, afin de décoller les lies de la paroi de la bouteille et de les faire descendre vers le goulot. (...)]₅₁₋₇₉
[Une fois le dégorgement effectué, dosée ou non, la bouteille de champagne sera bouchée avec son célèbre bouchon de liège maintenu par son muselet avant d’effectuer en cave un ultime vieillissement (maturation) avant commercialisation. On prélève sur ce stock, au fur et à mesure des besoins les bouteilles qui sont alors étiquetées et mises en caisses pour expédition.]₈₀₋₈₉

D’un côté, la perspective descendante met en évidence la structure énumérative annoncée par «consiste principalement à opérer une double fermentation du moût, la première en cuves, la seconde dans les bouteilles mêmes, en cave, avec remuage régulier» et deux items introduits par «la première fermentation» et «c’est cette deuxième fermentation».

[Vinification], [Le vin de Champagne est produit selon la méthode traditionnelle, jadis appelée méthode champenoise, qui consiste principalement à opérer une double fermentation du moût, la première en cuves, la seconde dans les bouteilles mêmes, en cave, avec remuage régulier. (...)] ₂₋₁₄	amorce
[La première fermentation, appelée fermentation alcoolique est identique à celle que subissent les vins tranquilles (c'est-à-dire non effervescents). Elle peut être suivie, mais ce n'est pas toujours le cas, d'une fermentation malolactique. (...)] ₁₅₋₃₀ [En début d'année (qui suit la récolte), les vins sont suffisamment clairs pour être goûtés et procéder à l'étape de l'assemblage qui mélange en proportions variant à chaque année des vins de cépages, terroirs et millésimes différents (aucune autre A.O.C. en France ne permet ce type de mélange de vin de différents millésimes).] ₃₁₋₃₈ [Quoi qu'il en soit, au moment d'embouteiller le vin de base ainsi obtenu, on lui ajoute la liqueur de tirage, composée de levures et de sucre. Cela enclenchera la dernière fermentation, dite prise de mousse.] ₃₉₋₄₄	item
C'est cette deuxième fermentation qui va donner naissance aux bulles de dioxyde de carbone. La bouteille est alors bouchée avec une capsule métallique analogue à celle des bouteilles de bière.] ₄₅₋₄₆ [Cependant, cette deuxième fermentation produit des lies abondantes dont on devra débarrasser le vin par la suite.] ₄₇₋₄₉	item
[Suit alors la période de vieillissement du vin en bouteilles d'une année environ pour les non millésimés à trois ans et plus pour les bouteilles millésimées.] ₅₀ [Après ce vieillissement, on rangeait autrefois les bouteilles sur des étagères appelées « pupitres » où elles étaient penchées le goulot vers le bas. Chaque jour, les bouteilles étaient remuées, c'est-à-dire tournées d'un quart de tour, d'un mouvement sec, afin de décoller les lies de la paroi de la bouteille et de les faire descendre vers le goulot. (...)] ₅₁₋₇₉ [Une fois le dégorgement effectué, dosée ou non, la bouteille de champagne sera bouchée avec son célèbre bouchon de liège maintenu par son muselet avant d'effectuer en cave un ultime vieillissement (maturation) avant commercialisation. On prélève sur ce stock, au fur et à mesure des besoins les bouteilles qui sont alors étiquetées et mises en caisses pour expédition.] ₈₀₋₈₉	

FIG. 9.13 – Annotation descendante de (29)

De l’autre côté, la perspective descendante met en évidence une structure plus large basée sur toutes les étapes de la vinification.

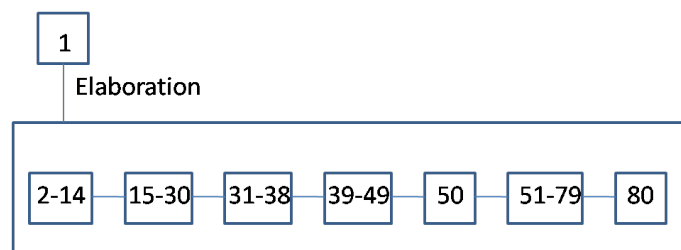


FIG. 9.14 – Annotation ascendante de nos deux annotateurs naïfs de (29)

De notre point de vue, ce texte présente une structure énumérative à deux temps, qui n'a été annoté ni par les annotateurs ascendants, ni par les annotateurs descendants. Les annotateurs ascendants ont mis en évidence la première structure présente dans les segments <1> à <6> tandis que les annotateurs descendants ont mis en évidence le second temps de la structure.

Commençons par les segments <1> à <6> mis en évidence par les annotateurs ascendants :

- (30) [Vinification,]₁ [Le vin de Champagne est produit selon la méthode traditionnelle,]₂
[jadis appelée méthode champenoise,]₃ [qui consiste principalement à opérer une
double fermentation du moût,]₄ [la première en cuves,]₅ [la seconde dans les bou-
teilles mêmes,]₆

Le segment <1> introduit le topique global du texte : la «vinification». Le segment <2> élabore ce segment : «le vin est produit selon la méthode traditionnelle». Une entité de ce segment «méthode traditionnelle» est élaborée dans le segment <3> : «méthode champenoise». Le segment <4> élabore l'éventualité de «méthode» en <2> qui consiste en une «double fermentation». Les segments <5> et <6> élaborent les deux étapes de la fermentation, tel qu'illustré Figure en 9.15.

Les segments <15-30>, <31-38> et <39-42>, mis au même niveau par les ascendants (Cf. Figure 9.14) que les segments <2-14>, décrivent, en trois étapes (l'étape de la «fermentation» en <15>, l'étape de l'«assemblage du vin» en <31> et l'étape de la «mise en bouteilles» en <39>), la première étape de la fermentation et sont donc dominés par le segment <4>, suivant nos propositions faites dans le chapitre 6. Toutes ces étapes sont accomplies avant celle de la «deuxième fermentation» en <43>, qui résulte de la mise en bouteille. On obtient ainsi une structure, illustrée en 9.15.

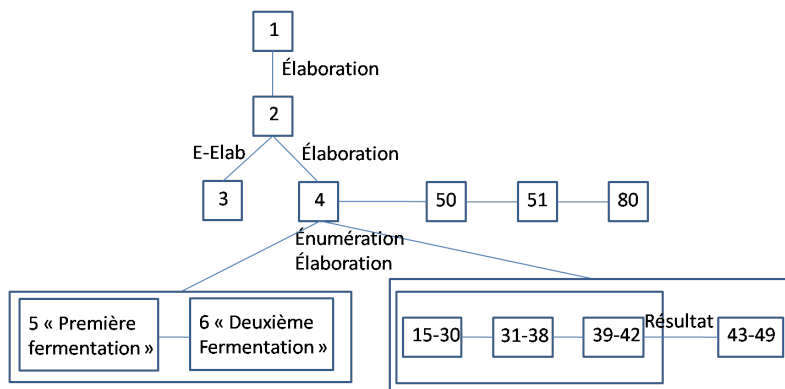


FIG. 9.15 – Annotation descendante de (29)

Contrairement à la structure énumérative à deux temps en (27), il n’y a pas de reprise anaphorique, du moins pour le second temps. La présentation de la deuxième fermentation est, en quelque sorte imbriquée dans la première, puisqu’elle en résulte : «cela enclenchera la dernière fermentation». Les deux temps de la structure énumérative sont ainsi moins mis en évidence.

Concernant les dernières étapes de la vinification décrites dans les segments <50> (la période de vieillissement), <51-79> (la période de repos) et <80-83> (la mise du bouchon) ne sont pas signalés par des marqueurs d’item, ce qui peut expliquer que les annotateurs descendants ne les aient pas signalés comme des items décrivant les étapes de la vinification. Les annotateurs ascendants, plus sensibles au contenu sémantique des parties du discours, ont annoté sous forme de blocs les segments correspondants aux différentes étapes de la vinification. Nous pouvons penser que l’introduction de la relation structurelle *Énumération* dans l’ensemble des relations de discours sensibiliserait les annotateurs à la reconnaissance dans les textes de structures signalées à la surface des textes et les aider à mettre en évidence la bonne structure du discours.

9.2.2.3 Bilan

Les structures énumératives organisées selon des marques de surface (lexicales et/ou typo-dispositionnelles) ont été plus facilement repérées et annotées par les annotateurs ascendants et descendants. Les deux modes d’organisation peuvent être rapprochés et méritent d’être combinés : par exemple l’approche ascendante détecte la hiérarchie présente dans les structures énumératives à deux temps que l’approche descendante présente de façon linéaire.

12 structures énumératives ont pu ainsi être rapprochées. Le rapprochement se fait sur la reconnaissance, par les annotateurs ascendants, de segments complexes correspondant aux items de l’approche descendante et d’une amorce dominant ces segments complexes. Dans ces 12 structures, les annotateurs s’accordent pour établir une relation d’*Élaboration* entre l’amorce et l’énumération, excepté dans deux cas (divergence avec *Explication* dans l’une et *Élaboration d’entité* dans l’autre ; l’annotation standard propose pour ces deux cas la relation d’*Élaboration*). Nous repérons dans ces structures 8 cas d’*Élaboration* clas-

siques. Enfin, dans l'exemple tel que le (26) avec une amorce incomplète (4 cas), nous proposons d'inférer la relation d'*Énumération* dans un premier temps, préalable nécessaire pour pouvoir insérer *Élaboration* dans un second temps.

Dans les 4 structures énumératives restantes, les modes d'organisation ont mis en avant des organisations différentes de la structure du texte. Nous rejetons le statut de structure énumérative pour l'une d'elles (28). Une seconde ne contient pas d'amorce. Les deux autres sont traitées au moyen de la relation d'*Élaboration*.

L'analyse des annotations de la perspective ascendante des structures énumératives annotées dans la perspective descendante nous permet de conclure 1. qu'introduire la relation d'*Énumération* dans l'ensemble des relations fournies aux annotateurs aurait pu les aider à représenter plus efficacement des structures complexes et 2. que les structures énumératives et les indices permettant de les repérer constituent un vivier pour l'identification de structures élaboratives.

9.2.3 Via l'annotation standard

C'est dans cette section que nous abordons de front une perspective onomasiologique. Comme nous ne pouvions pas, dans le temps qui nous était imparti, regarder les 1825 annotations pour les deux relations d'élaboration, nous nous sommes concentrés sur la relation d'*Élaboration* et sur 244 paires d'annotations extraites de 16 textes doublement annotés. 11 de ces textes ont fait l'objet d'une annotation standard proposée pour chaque texte par deux membres du projet après concertation. Les 5 autres textes ont été choisis aléatoirement et nous proposons pour ces textes une annotation standard pour les élaborations.

Double annotation	Nombre de textes
Sophie/Nadej	10
Simon/Sophie	3
Simon/Nadej	3

Notre programme permet de trier dans les annotations de chaque texte les configurations suivantes :

AA-AR (Accord Attachement - Accord Relation)	84 cas
AA-DR (Accord Attachement - Désaccord Relation)	146 cas
DA (Désaccord Attachement)	11 cas

Nous allons, dans les sections suivantes, proposer une évaluation qualitative de ces doubles annotations, en gardant à l'esprit notre objectif, i.e. mettre au jour les réalisations linguistiques de la relation d'*Élaboration*.

9.2.3.1 Analyse des accords entre annotateurs (AA-AR)

- *Élaboration* dans le standard (32 cas)

Nous allons considérer ces 32 cas pour mettre au jour des indices de la relation d'*Élaboration*. Commençons avec l'exemple (31).

- (31) [Tonneaux à Montreux-Vieux :]₁ [trois blessés graves]₂
 [Trois personnes ont été gravement blessées,]₃ [dimanche soir vers 17 h 40 à la suite
 d'un accident spectaculaire]₄ [qui s'est produit à l'entrée du village de Montreux-
 Vieux sur la départementale 32]₅
AA-AR : Élaboration (2,3)

Les segments <2> et <3> sont reliés par la relation d'*Élaboration*. Le segment <3>, premier segment du paragraphe, reprend les mots du titre. Les reprises lexicales peuvent être considérées comme des indices sur lesquels s'appuie l'interprétation d'une élaboration.

Comme attendu, peu d'exemples font apparaître un marqueur lexical en position initiale, comme *c'est-à-dire* en (32).

- (32) [D'un point de vue technique,]₆ [BITNET était différent d'Internet]₇ [parce que
 c'était un réseau point-à-point de type «stocké puis transmis»].₈ [**C'est-à-dire** que
 les messages e-mail et les autres fichiers étaient transmis en entier d'un serveur à
 l'autre jusqu'à leur destination finale.]₉
AA-AR : Élaboration (8,9)

C'est-à-dire introduit des reformulations (Vassiliadou, 2004), il apparaît donc comme un candidat idéal pour l'identification de la relation d'*Élaboration*. Mais un bref retour sur les occurrences de ce marqueur dans les annotations signale qu'il n'introduit pas toujours une élaboration sur une éventualité. Son rôle peut également être d'introduire une élaboration sur une entité.

L'introduction d'une éventualité plurielle, comme en (33) et (34), peut dans certains cas préparer le lecteur à interpréter une élaboration.

- (33) [C'est connu,]₂₃ [les malheurs ne surviennent jamais seuls :]₂₄ [un important groupe
 canadien [qui se disait intéressé par la distribution en Europe des plaques numé-
 riques mises au point à Baume]₂₆ n'a soudainement plus donné signe de vie.]₂₅
 [Quant au chèque de 76.000 euros,]₂₇ [la banque a apparemment refusé de l'honorer.]₂₈

La forme plurielle de «malheurs» accompagnée de l'utilisation du signe ponctuation « : » prépare le lecteur à l'élaboration. On s'attend en effet à interpréter deux événements qui sont des malheurs, respectivement introduits par les segments (25) et (27-28).

- (34) [Les attentats du 11 septembre 2001 sont une série d'attentats-suicides commis par
 le réseau terroriste Al-Qaida dans le nord-est des États-Unis le mardi 11 septembre
 2001.]₂ [Quatre avions de ligne sont détournés,]₄ [trois d'entre eux sont projetés
 contre des immeubles hautement symboliques :]₅ [les tours jumelles du World Trade
 Center à Manhattan, [à New York,]₇ et le Pentagone,]₆ [siège du département de
 la Défense des États-Unis,]₈ [à Washington.]₉ [Les tours se sont effondrées moins
 de deux heures plus tard entraînant l'immeuble du Marriott World Trade Center
 dans leur chute.]₁₀ [La tour 7 du WTC s'est effondrée dans l'après-midi]₁₁ [en
 raison d'incendies et des dégâts occasionnés par la chute des Twin Towers.]₁₂ [Le

quatrième avion s'est écrasé en rase campagne à Shanksville,]₁₃ [en Pennsylvanie.]₁₄
AA-AR : Élaboration (2,4) & Élaboration (4,[5-14])

La première élaboration de cet extrait (*Élaboration(2,4)*) est en quelque sorte annoncée dans le segment <2> «les attentats sont une **série** d'attentats-suicides». Le segment <4> annonce alors le nombre d'éléments de cette série s'élevant à 4 «**Quatre** avions de ligne sont détournés». La deuxième élaboration (*Élaboration(4,[5-14])*) s'instaure dans la continuité : <2> ayant annoncé qu'il y a eu une série d'attentats, puis <4> ayant spécifié que cette série s'élève à 4 attentats, le lecteur est dans l'attente d'une élaboration de ces 4 attentats. Cette attente est satisfaite avec les segments (5) et (13) : «**trois d'entre eux** sont projetés contre des immeubles hautement symboliques» et «**Le quatrième avion** s'est écrasé en rase campagne». La structure de ce discours est donnée Figure 9.16.

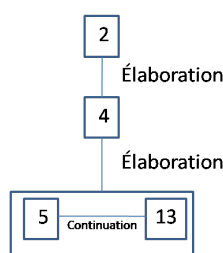


FIG. 9.16 – Structure de (34)

Les indices de cette élaboration sont là aussi l'introduction d'un événement pluriel, accompagnée de la mention d'adjectifs numéraux. Bien évidemment, ces indices ne sont pas suffisants à eux-seuls pour identifier des élaborations dans un texte mais ils pourront certainement s'avérer utiles s'ils sont combinés à d'autres indices.

Comme nous l'avons vu dans la section 9.2.2, des indices de surface (lexicaux et typographiques) facilitent la reconnaissance d'une structure énumérative.

- (35) [Les serments de Strasbourg [(Sacramenta Argentaria)]₃ sont un texte de 842,]₂
 [écrit en trois langues :]₄
 [- Dans une langue romane composite]₅ [qu'on estime être du proto-français]₆ [(ou
 [selon d'autres hypothèses,]₈ du proto-occitan,]₇ [ou un mélange voulu de français
 et d'occitan),]₉
 [- En proto-allemand]₁₀
 [- En latin.]₁₁
AA-AR : Élaboration (4,[5-11])

Dans le cas d'une structure énumérative comme en (35), la structure énumérative (reconnaissable à l'emploi d'un prospect «en trois langues», du signe de ponctuation « : » et des tirets) se superpose à une structure élaborative.

En (36), on reconnaît une amorce annonçant une énumération. Le contenu événementiel du SN prospectif «deux phases» annonce, quant à lui, une élaboration, i.e. le développement

de l'événement en deux phases du «réchauffement» et les marqueurs *le premier* et *le second* introduisent les segments élaborants :

- (36) [Ce réchauffement se serait déroulé en deux phases,]₄₂ [la première de 1910 à 1945,]₄₃ [la seconde de 1976 à aujourd'hui.]₄₄
AA-AR : Élaboration (42,[43-44])

Les relatives permettent d'introduire des élaborations, comme la relative introduite par le pronom relatif *où* en (37)

- (37) [Elle tourne [(2006)]₈₅ un nouveau film avec Nadir Moknèche]₈₄ [**où** elle joue une mafieuse]₈₆
AA-AR : Élaboration (84,86)

La relative introduit l'événement «jouer le rôle d'une mafieuse» qui élabore l'événement de la principale «tourner dans un film».

On trouve également des relatives introduites par le pronom relatif *qui* comme en (38).

- (38) [choisissant avec soin un répertoire franco-algérien]₇₂ [**qui** puise dans les deux cultures.]₇₃
AA-AR : Élaboration (72,73)

Enfin, nous relevons des relatives introduites par le pronom relatif *dont* comme en (39).

- (39) [L'élévation de la température moyenne du globe entre 1906 et 2005 est estimée à 0,74 °C]₄₉ [(à plus ou moins 0,18 °C près),]₅₀ [**dont** une élévation de 0,65 °C durant la seule période 1956-2006]₅₁
AA-AR : Élaboration (50,51)

Les relatives apparaissent donc comme des vecteurs de la relation d'*Élaboration*. Ce n'est en revanche pas des marqueurs de la relation comme en attestent les nombreux exemples de relatives pour la relation d'*Élaboration d'entité* dans le manuel.

En (40), l'élaboration est introduite par la préposition *avec*.

- (40) [Du 10 au 21 janvier,]₃₀ [la température sous-abri se maintint à Paris aux environs de -20 °C,]₃₁ [**avec** des minima absolus de -23 °C les 13 et 14 janvier;]₃₂
AA-AR : Élaboration (31,32)

Pour inférer l'élaboration en (40), il faut comprendre «des minima absolus de -23 °C comme précisant «aux environs de -20»²

•Élaboration d'entité dans le standard (23 cas)

²Nous notons qu'avec *en dessous de*, l'élaboration serait encore plus claire : [Du 10 au 21 janvier,]₃₀ [la température sous-abri se maintint à Paris aux environs de -20 °C,]₃₁ [**avec** des minima absolus de -23 °C les 13 et 14 janvier;]₃₂.

Annotation standard	Nombre de cas
Élaboration d'entité	27
Élaboration	10
Fusion	5
Arrière-plan	2
Flashback	2
Explication	2
Attribution	1
Commentaire	1
Narration	1
Total	51

TAB. 9.14 – Annotation standard dans les cas de divergence *Élaboration* vs *Élaboration d'entité*

L'exemple suivant illustre un cas fréquent où les deux annotateurs ont privilégié la relation d'*Élaboration* tandis que nous proposons la relation d'*Élaboration d'entité* entre les segments (53) et (54).

- (41) [MAD était écrit en REXX,]₅₂ [et proposait plusieurs labyrinthes multi-étages,]₅₃
 [peuplés de robots (...) mobiles communicants]₅₄
AA-AR : Élaboration(53,54)

Ce premier cas de divergence entre *Élaboration* et *Élaboration d'entité* peut être, de notre point de vue, facilement évité. La relation d'*Élaboration d'entité* porte sur une entité «labyrinthe» du segment <52>.

9.2.3.2 Analyse des divergences (AA-DR)

Divergence entre *Élaboration* et *Élaboration d'entité*

Pour rappel, la relation d'*Élaboration d'entité* a été introduite dans (Prévot *et al.*, 2009) après un constat de divergence très fréquente entre les relations d'*Arrière-plan* et d'*Élaboration* (Cf. chapitre 3. Dans le manuel d'annotation, cette relation est décrite comme reliant deux segments si le segment à attacher précise une propriété d'une des entités impliquées dans le segment-cible.

En comparant toutes les annotations des annotateurs naïfs, nous avons relevé un fort taux de divergence entre les relations d'*Élaboration* et d'*Élaboration d'entité*. 40% des divergences d'*Élaboration* se font avec *Élaboration d'entité* et 70% des divergences d'*Élaboration d'entité* se font avec *Élaboration*.

Dans l'échantillon des 16 textes sélectionnés, nous relevons 51 paires d'annotations divergentes *Élaboration* vs. *Élaboration d'entité*. L'annotation standard opte le plus souvent pour l'une (*Élaboration* dans 20% des cas) ou l'autre (*Élaboration d'entité* dans 53 % des cas) des relations, comme nous le voyons dans le tableau 9.14.

• *Élaboration d'entité* dans l'annotation standard

Le premier exemple de divergence que nous allons considérer est un cas typique de la relation *Élaboration d'entité*. Les segments <25-26-27> apportent des propriétés, tantôt définitoires, tantôt marginales, d'une entité **secondaire** «arme de poing» introduite dans le segment <24>.

- (42) [Il aurait pu se contenter de venir les mains vides,]₂₃ [mais il est arrivé avec une arme de poing,]₂₄ [Vraisemblablement un revolver,]₂₅ [calibre 22 LR,]₂₆ [qu'il a caché à Murielle en le tenant dans son dos.]₂₇
Annot1 : *Élaboration* (24,25)
Annot2 : *Élaboration d'entité* (24,25)

La divergence avec la relation d'*Élaboration* ne semble se justifier que par une mauvaise compréhension de la relation d'*Élaboration d'entité*.

Les titres sont difficiles à annoter. Ils montrent, en quelque sorte, une certaine résistance à l'analyse en relations de discours. Ce qui est certain, c'est qu'une seule relation de discours ne peut pas permettre de rendre compte du rôle discursif de tous les titres.

- (43) [Le piano moderne]₁
 [Le clavier]₂
 [Le clavier du piano est composé de 88 touches (sauf exception).]₃ [Les 52 touches blanches correspondent aux notes non altérées,]₄ [et les 36 touches noires aux notes altérées (dièse ou bémol)]₅.
Annot1 : *Élaboration* (1,2) & *Élaboration*(2,[3,8,22,30,33])
Annot2 : *Élaboration d'entité* (1,2) & *Élaboration d'entité* (2,[3,8,22,30])

La relation d'*Élaboration d'entité* semble être l'annotation la plus appropriée : le segment <2> apporte des informations sur une entité impliquée dans le segment <1>, de même les segments <3> et suivants apportent des informations sur une entité impliquée dans le segment <2>. Il nous semble, par ailleurs, que la description de la relation d'*Élaboration* pourrait convenir également du fait qu'une seule entité est impliquée dans les segments <1> et <2>. Les deux annotations nous paraissent similaires dans ce cas.

• *Élaboration* dans l'annotation standard

- (44) [Attentats du 11 septembre 2001.]₁
 [Les attentats du 11 septembre 2001 [(abréviations : 11/9, 11-Septembre et, en anglais, 9/11)]₃ sont une série d'attentats-suicides commis par le réseau terroriste Al-Qaïda dans le nord-est des États-Unis le mardi 11 septembre 2001.]₂
Annot1 : *Élaboration* (1,2)
Annot2 : *Élaboration d'entité* (1,2)

Contrairement au titre de (43), le titre en (44) décrit un événement. Le segment <2> décrit cet événement. La relation entre les segments <1> et <2> est donc *Élaboration*.

Dans peu de cas, une autre relation est proposée par l'annotation standard, comme en (45) analysé au moyen de la relation de *Flashback* ou en (46) analysé au moyen de la relation d'*Explication*.

• *Flashback* dans l'annotation standard

- (45) [Un important groupe canadien [qui se disait intéressé par la distribution en Europe des plaques numériques mises au point à Baume]₂₆ n'a soudainement plus donné signe de vie.]₂₅

Annot1 : Élaboration (25,26)

Annot2 : Élaboration d'entité (25,26)

«Se dire intéressé par...» dans le segment <26> n'est pas un sous-événement de l'événement «ne donner plus signe de vie» du segment <25>. En revanche, la relation d'*Élaboration d'entité* semble se justifier suivant les indications données dans le manuel. La relative en <26> caractérise effectivement une entité du segment <25>, i.e. le GN «groupe canadien», en donnant une propriété de celui «se dire intéressé par la distribution en Europe...». Pourtant une analyse plus fine du rôle discursif du segment <26> montre qu'il décrit un état se situant avant l'état décrit en <25>. La relation de *Flashback* est donc préférée.

• *Explication* dans l'annotation standard

De même dans l'exemple (46), un annotateur a relié le segment <9> au segment <8> par *Élaboration* tandis que l'autre annotateur a utilisé la relation *Élaboration d'entité*.

- (46) [Les touches du piano sont généralement faites en épicéa ou en tilleul,]₈ [bois choisis pour leur légèreté.]₉

Annot1 : Élaboration (8,9)

Annot2 : Élaboration d'entité (8,9)

Dans ce cas aussi, la relation d'*Élaboration d'entité* semble se justifier, le segment <9> donne des caractéristiques des entités «épicéa ou tilleul». Or le segment <9> peut également être interprété comme une explication : «les touches du piano sont faites en épicéa ou en tilleul parce que ce sont des bois légers».

Notre conclusion vis-à-vis de ces divergences est qu'il faut proposer aux annotateurs un ordre de priorité d'application des relations de discours, i.e. annoter *Élaboration d'entité* si aucune autre relation du modèle d'annotation ne convient.

Divergence entre *Élaboration* et *Encadrement* La relation d'*Encadrement* a été introduite spécialement pour permettre aux annotateurs de marquer la relation d'indexation (Charolles, 1997) établie entre l'adverbial détaché et les segments qu'il domine.

Bien que cette consigne d'annotation n'apparaisse pas difficile à appliquer, on relève 17 paires d'annotations divergentes *Élaboration* vs. *Encadrement* dans les 16 textes sélectionnés. L'annotation standard opte le plus souvent pour la relation d'*Encadrement* (dans 60% des cas) ou la relation de fusion (dans 30% des cas), comme nous le voyons dans le tableau 9.15.

Annotation standard	Nombre de cas
Encadrement	10
Fusion	5
Arrière-plan	1
Élaboration d'entité	1
Total	17

TAB. 9.15 – Annotation standard dans les cas de divergence *Élaboration* vs *Encadrement*

Parmi les adverbiaux détachés concernés par cette divergence d'annotation, nous distinguons trois types de cadres : les cadres temporels, les cadres thématiques et les cadres médiatifs.

•Cadres temporels

L'annotateur qui propose *Encadrement* se fie correctement au manuel tandis que l'annotateur qui propose *Élaboration* se fie à une intuition validant, en quelque sorte, les hypothèses avancées par Vieu *et al.* (2005). Celui-ci a eu l'intuition que les segments dominés par le cadre élaborent ce qui s'est passé dans la période définie par celui-ci.

- (47) [Cet**te** ann**ée** là,]₂₀ [les céréales manqu**è**rent dans la plus grande partie de la France,]₂₁ [et seuls la Normandie, le Perche et les côtes de Bretagne ont pu produire assez de grain]₂₂ [pour assurer les semences.]₂₃ [Dans la région parisienne]₂₄ [le prix du pain atteignit, [en juin 1709,]₂₆ 35 sous les neuf livres]₂₅ [au lieu de 7 sous ordinairement.]₂₇ [De nombreux arbres gel**è**rent jusqu'à l'aubier,]₂₈ [et la vigne disparut de plusieurs régions de la France.]₂₉ [Du 10 au 21 jan**vier**,]₃₀ [la température sous-abri se maintint à Paris aux environs de -20 ° C,]₃₁ [avec des minima absolus de -23 ° C les 13 et 14 jan**vier** ;]₃₂ [le 11,]₃₃ [le thermomètre s'abaissa jusqu'à -16 ° C à Montpellier et -17 ° C à Marseille.]₃₄
Annot1 : *Élaboration*(20,[21-29]) & *Élaboration*(30,[31-32]) & *Élaboration*(33,34)
Annot2 : *Encadrement* (20,[21-29]) & *Encadrement* (30,[31-32]) & *Encadrement* (33,34)

Ces annotations avec *Élaboration*, bien que contraires aux spécifications du manuel, nous paraissent très intéressantes, fournissant une validation à l'hypothèse de les traiter comme des introducteurs de nouveaux topiques.

•Cadres thématiques

Ces cadres de discours sont généralement introduits par des expressions topicalisatrices, comme illustré en (48) :

- (48) [Quant à la gendarmerie de Dannemarie,]₃₀ [elle a procédé au constat d'usage.]₃₁
Annot1 : *Élaboration*(30,31)
Annot2 : *Encadrement* (30,31)

Annotation standard	Nombre de cas
Continuattion	5
Élaboration	3
Contraste	1
Résultat	1
Total	10

TAB. 9.16 – Annotation standard dans les cas de divergence *Élaboration* vs *Continuation*

Ces cadres introduisent un référent à propos duquel on va dire quelque chose. Bien que contraire aux spécifications du manuel, la relation d'*Élaboration* nous apparaît comme tout à fait justifiée pour traiter ces cadres de discours qui introduisent un référent sur lequel on va élaborer.

• Cadres médiatifs

Les cadres médiatifs indexent des segments en les affectant à une source : «Selon la bible», «Pour Guiraud».

- (49) [Pour Guiraud,]₄₁ [il existe une fréquence intrinsèque des mots dans la langue,]₄₂
 [ce que conteste Mandelbrot.]₄₃
Annot1 : *Élaboration*(41,42)
Annot2 : *Encadrement* (41,42)

La relation d'*Encadrement* est appropriée pour traiter cet attachement selon les spécifications du manuel. *Élaboration*, en revanche, ne semble pas convenir. Il semblerait plus adéquat d'attacher les segments dominés par le cadre au moyen de la relation d'*Attribution*.

Pour notre recherche des marqueurs et indices de la relation d'*Élaboration*, ces divergences ont permis d'accéder à des données très intéressantes et des indices (cadres spatiaux et cadres thématiques) à considérer pour la recherche de structures élaboratives dans les textes. Le rôle discursif des cadres spatio-temporels a déjà été décrit dans la mise en œuvre de la relation d'*Élaboration*. Les cadres thématiques doivent, quant à eux, être étudiés dans cette perspective.

Divergence entre *Élaboration* et *Continuation* Dans l'échantillon des 16 textes sélectionnés, nous relevons 10 paires d'annotations divergentes *Élaboration* vs. *Continuation*. L'annotation standard opte le plus souvent pour la relation de *Continuation* (dans 50% des cas) ou d'*Élaboration* (dans 30 % des cas), comme nous le voyons dans le tableau 9.16.

Dans de nombreux cas, la divergence avec la relation de *Continuation* provient d'une double interprétation possible, illustrée en (50) :

- (50) [Elle parle des rapports amoureux,]₅₆ [fustige l'intolérance, les mesquineries et la bêtise,]₅₇
Annot1 : Élaboration(56,57)
Annot2 : Continuation (56,57)

Deux éventualités sont décrites «parler des rapports amoureux» et «fustiger l'intolérance, les mesquineries et la bêtise». On peut interpréter la deuxième éventualité comme «fustiger l'intolérance, les mesquineries et la bêtise dans les rapports amoureux», et donc comme une élaboration de la première éventualité. On peut aussi interpréter les deux éventualités comme indépendantes et donc proposer *Continuation*.

• *Continuation* dans l'annotation standard

On comprend assez facilement les divergences entre *Élaboration* et *Continuation* observées dans les exemples (51) et (52).

- (51) [De plus, de nombreux courtiers en déchets cherchèrent à exploiter l'image de marque du recyclage]₁₃ [et commencèrent à présenter les destinations de leurs exportations comme des sites de recyclage.]₁₄
Annot1 : Élaboration (13,14)
Annot2 : Continuation (13,14)
- (52) [Biyouna continue une carrière de chanteuse]₆₅ [et sort en 2001 un album de Chanson Raid Zone,]₆₆
Annot1 : Élaboration (65,66)
Annot2 : Continuation (65,66)

On peut comprendre, en (51), que le type de l'éventualité «commencer à présenter les destinations...» est un sous-type ou une façon de «chercher à exploiter l'image de marque». De même, en (52), l'éventualité «sortir un album» peut être comprise comme faisant parti de l'éventualité «continuer une carrière de chanteuse». Mais, de notre point de vue, ce ne sont pas des cas classiques d'*Élaboration*. La conjonction de coordination *et* oriente l'interprétation vers l'inférence de la relation de *Continuation*. Les premiers événements «chercher à exploiter l'image de marque» et «continuer une carrière de chanteuse» ont, tous les deux, une lecture inchoative et la borne initiale des événements «commencer à présenter les destinations» et «sortir un album» se situe nécessairement après celle des événements décrits dans le premier segment.

• *Élaboration* dans l'annotation standard

Dans certains cas, comme illustré en (53), la divergence semble liée à des erreurs d'interprétation (mais qui ne le sont certainement pas dans la tête de l'annotateur).

- (53) [C'est dans le Nouveau Testament que l'on trouve trace de Capharnaüm pour la première fois.]₁₃ [Cette ville est abondamment citée dans les récits de l'Évangile]₁₄
Annot1 : Élaboration (13,14)
Annot2 : Continuation (13,14)

Annotation standard	Nombre de cas
Élaboration	3
Explication	2
Résultat	2
Élaboration d'entité	2
Arrière-plan	1
Attachement non pertinent	1
Total	11

TAB. 9.17 – Annotation standard dans les cas de divergence *Élaboration* vs *Explication*

Pour interpréter correctement l'attachement des segments <13> et <14>, il faut savoir que les récits de l'Évangile sont une partie du Nouveau Testament. Ainsi, on peut attacher le segment <14> au segment <13> par la relation d'*Élaboration*. Si l'on ne connaît pas cette information et que l'on considère le Nouveau Testament et les récits de l'Évangile comme deux ouvrages distincts, alors l'annotation de la relation de *Continuation* est compréhensible.

Divergence entre *Élaboration* et *Explication*

Dans l'échantillon des 16 textes sélectionnés, nous relevons 11 paires d'annotations divergentes *Élaboration* vs. *Explication*. La divergence totale avec l'annotation standard semble indiquer les difficultés rencontrées par les annotateurs concernant les deux relations d'*Élaboration* et *Explication*, comme nous le voyons dans le tableau 9.17.

• *Élaboration* dans l'annotation standard

- (54) [À l'intérieur des grandes fluctuations climatiques terrestres,]₂ [se trouvent des variations plus brèves et plus limitées en intensité.]₃ [**Ainsi**,]₄ [au cours du dernier millénaire,]₅ [est apparu une période chaude aux xe et xie siècles]₆ [appelée «optimum climatique médiéval»]₈ :]₇ [De même,]₁₃ [l'époque des Temps Modernes [(1550-1850)]₁₅ connu une période de refroidissement]₁₄

Annot1 : *Élaboration* (3,4)

Annot2 : *Explication* (3,4)

En (54), le marqueur *ainsi* en position initiale introduit des exemples de variations climatiques. L'annotation standard propose *Élaboration*.

Le marqueur *ainsi* en position initiale est relevé dans un autre segment élaborant de notre corpus. Un bref retour sur les annotations permet de montrer que ce marqueur n'apparaît pas que dans des élaborations, mais aussi pour introduire des continuations ou des résultats comme l'a montré (Atallah, 2010) dans son exploration des relations causales dans le corpus ANNODIS. Dans 6 cas sur 8, nous relevons des annotations divergentes du segment contenant *ainsi* au contexte gauche. Ces indications sont révélatrices des difficultés quant aux indications qu'apporte ce marqueur pour l'annotation en relations de discours.

Dans le chapitre 7, nous avons signalé que les participes présents peuvent être des élaborations de la proposition principale comme en (55).

- (55) [Biyouna a pris son temps,]₇₁ [**choisissant** avec soin un répertoire franco-algérien]₇₂
Annot1 : Élaboration (71,72)
Annot2 : Explication (71,72)

En (55), un annotateur a proposé la relation d'*Explication*, considérant «choisir avec soin un répertoire» comme expliquant pourquoi Biyouna a pris son temps. De notre point de vue, le participe présent est un indice jouant en faveur de la relation d'*Élaboration*, les deux procès étant présentés comme se déroulant simultanément.

• *Explication* dans l'annotation standard

Dans les exemples (56) et (57), il n'est pas évident de déterminer les raisons qui ont poussé les annotateurs à choisir la relation d'*Élaboration*.

En (56), le marqueur *en effet* introduit une justification ou une explication du «mauvais choix». On comprend que les Khazars avaient protégé les Russes et maintenant que leur empire n'existe plus, ces derniers ne les protègent plus, expliquant ainsi que ce fut un mauvais choix pour les Russes.

- (56) [Finalement,]₁₈ [la fin de l'Empire khazar s'avéra un mauvais choix politique pour les Russes :]₁₉ [les Khazars, en effet, les avaient protégés contre les Petchenègues]₂₀ [qui nomadisaient au sud de la Russie.]₂₂
Annot1 : Élaboration(19,20)
Annot2 : Explication (19,20)

En (57), le segment (4) introduit un événement de type «accident» qui peut être une explication du type de l'état décrit dans le segment (3) «trois personnes ont été gravement blessées».

- (57) [Trois personnes ont été gravement blessées,]₃ [dimanche soir vers 17 h 40 à la suite d'un accident spectaculaire]₄
Annot1 : Élaboration(3,4)
Annot2 : Explication (3,4)

Certains cas de divergence résultent de réelles ambiguïtés interprétatives.

Mais dans la majorité des cas de divergences, une seule interprétation n'est possible. Ces divergences auraient, nous semble-t-il, pu être évitées à condition que les annotateurs aient été plus longuement formés et entraînés.

Enfin, l'analyse de ces divergences permet de relever des cas intéressants pour l'analyste, des cas où la tension entre deux relations doit être résolue par l'analyste expert pour aider l'annotateur dans sa tâche d'annotation.

L'analyse de ces divergences permet également de relever des indices de la relation. Le fait que ces indices apparaissent dans des annotations divergentes laissent attendre qu'ils sont plutôt des indices ambigus. Par exemple, les titres peuvent introduire, dans la structure, un constituant qui va être par la suite élaboré. La nature du titre va être

déterminante pour la nature de la relation inférée. Si le titre introduit un référent, ce dernier sera sans doute attaché au contexte droit par la relation d'*Élaboration d'entité*. En revanche, s'il introduit un événement, ce dernier sera sans doute attaché au contexte droit par la relation d'*Élaboration*.

9.2.3.3 Analyse des désaccords (DA)

Pour terminer notre exploitation des annotations d'*Élaboration*. Nous relevons des cas où seul un annotateur a proposé un attachement entre deux segments avec la relation d'*Élaboration*. L'autre annotateur n'a tout simplement pas attaché ces segments là. Nous appelons les attachements concernés *Élaboration longue distance*, aussi appelés *Discourse pop* en anglais. Ces attachements font intervenir deux segments qui ne sont pas côte à côte dans le texte.

Dans certains cas d'*Élaboration longue distance*, on note que la cohésion est assurée, entre autres, par un mécanisme de répétition lexicale ou syntagmatique.

- (58) [Trois personnes **ont été gravement blessées**,]₃ [dimanche soir vers 17 h 40 à la suite d'un accident spectaculaire]₄ [qui s'est produit à l'entrée du village de Montreux-Vieux sur la départementale 32]₅ (...) [Le conducteur, [Benattia Grabis,]₁₅ (...) [ainsi que ses deux occupants,]₁₈ [Abdel Rhazza,]₁₉ (...) [et Lionel Ledoucin,]₂₂ **ont été gravement blessés** dans l'accident.]₁₄
Annot1 : Élaboration (3,14)
- (59) [Une pluie d'**étoiles**]₁
[Deux étoiles] pour l'hôtel Le Relais,]₁₂ [situé sur la place de l'Eglise;]₁₃ **[une étoile]** pour le restaurant du même nom,]₁₄ [bien connu des habitants et des touristes.]₁₅ (...) **[Deux étoiles]** aussi,]₁₉ [c'est le nouveau classement de l'Office de tourisme d'Arc-et-Senans]₂₀ [qui prend ainsi sa place dans la cour des grands,]₂₁ [juste derrière les offices de Besançon et Montbéliard.]₂₂ *Annot1 : Élaboration (1,[12-22])*
- (60) [un Nancéen **tire plusieurs fois** sur sa compagne]₃ [qu'il blesse grièvement.]₄ (É)
 [La rencontre se passe mal]₂₈ [et Michel empoigne son ancienne compagne]₂₉ [et **tire plusieurs fois** sur elle.]₃₀ [Une balle la touche à la cuisse]₃₁ [et se loge derrière le genou,]₃₂ [une autre l'atteint à la hanche,]₃₃ [une troisième lui arrive dans le front.]₃₄
Annot1 : Élaboration (3,30)

En (58), le SV «ont été gravement blessées» est repris. En (59), le nom «étoiles» est répété deux fois. Et en (60), le SV «tire plusieurs fois» est répété.

En (61), une *Élaboration longue distance* est ici supportée par des cadres de discours qui organisent thématiquement le passage. Ces cadres font référence à des périodes temporelles imbriquées les unes dans les autres :

- (61) [C**ette année là**,]₂₀ [les céréales manquèrent dans la plus grande partie de la France,]₂₁ [et seuls la Normandie, le Perche et les côtes de Bretagne ont pu produire assez de grain]₂₂ [pour assurer les semences.]₂₃ [Dans la région parisienne]₂₄ [le prix du pain atteignit, [en juin 1709,]₂₆ 35 sous les neuf livres]₂₅ [au lieu de 7 sous ordinairement.]₂₇ [De nombreux arbres gelèrent jusqu'à l'aubier,]₂₈ [et la vigne disparut de plusieurs régions de la France.]₂₉ [D**u 10 au 21 janvier**,]₃₀ [la température sous-abri se maintint à Paris aux environs de -20 ° C,]₃₁ [avec des minima absolus de -23 ° C les 13 et 14 janvier ;]₃₂ [l**e 11**,]₃₃ [le thermomètre s'abaissa jusqu'à -16 ° C à Montpellier et -17 ° C à Marseille.]₃₄
Annot1 : Élaboration (20,30) & Élaboration (30,33)

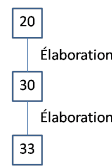


FIG. 9.17 – Structure de (61)

9.2.3.4 Rappel et précision

En proposant la confrontation des annotations des annotateurs «naïfs» à l'annotation standard, il nous est possible de donner les mesures de précision et de rappel dans le tableau 9.18. La précision est le calcul permettant de mettre en valeur, pour un annotateur, le nombre d'annotations d'*Élaboration* qui sont validées par l'annotation standard. Le rappel est le calcul permettant de mettre en valeur le nombre d'annotations d'*Élaboration* du standard que l'on retrouve dans l'annotation d'un annotateur donné. La F-mesure est le calcul permettant de pondérer les calculs de rappel et de précision, particulièrement lorsque ces derniers sont déséquilibrés (précision forte et rappel mauvais ou l'inverse).

Annotateurs	Nombre <i>Élaboration</i> par annotateur	Nombre <i>Élaboration</i> dans l'annotation standard	<i>Élaboration</i> commune	Précision	Rappel	F-Mesure
Simon	34	24	14	0,4	0,6	0,48
Nadej	183	64	183	0,3	0,9	0,45
Sophie	174	68	41	0,2	0,6	0,3

TAB. 9.18 – Précision, Rappel et F-Mesure par annotateur

La précision pour chaque annotateur est très basse. Elle est définie par le nombre d'*Élaboration* annotées communes rapporté au nombre d'*Élaboration* de l'annotateur. Les annotateurs «naïfs» annotent beaucoup plus d'*Élaboration* que les annotateurs experts.

Le rappel est meilleur. Il est défini par le nombre d'*Élaboration* annotées communes rapporté au nombre d'*Élaboration* du standard. On note un fort rappel pour Nadej. Cette

annotatrice a annoté quasiment toutes les élaborations du standard. Mais ce bon résultat doit être pondéré par le fait que cette annotatrice a annoté beaucoup plus d'*Élaboration* que les annotateurs experts (précision très faible), ce que la F-mesure nous permet de montrer.

Les scores de précision et de rappel sont mauvais. De plus lorsque les annotateurs s'accordent sur la relation d'*Élaboration*, ils ne s'accordent que dans 43% des cas avec l'annotation standard. En cas d'annotations divergentes, l'annotation standard tranche dans seulement 20% des cas pour la relation d'*Élaboration*, dans 45% des cas pour l'annotation divergente et dans 35% des cas pour une autre relation de discours.

Sur les 244 paires, seuls 66 attachements ont été annotés comme des cas d'*Élaboration* dans le standard, soit moins du tiers. 62 ont été annotés avec *Élaborations d'entité*. 31 attachements ont été fusionnés, montrant que la relation d'*Élaboration* semble être privilégiée par les annotateurs naïfs dans les cas où il n'y a pas de relation du tout ! Puis on retrouve, dans une moindre mesure, les divergences avec les relations d'*Arrière-plan*, d'*Encadrement*, de *Résultat*, de *Commentaire*...

9.2.3.5 Bilan et perspectives

Nous avons adopté une démarche onomasiologique avec l'objectif de mettre au jour des réalisations linguistiques de la relation d'*Élaboration*. Mais cette démarche ne peut suffire pour considérer un indice comme un marqueur de la relation pour deux raisons principales, que nous avons déjà abordées, la faible quantité et la faible qualité des annotations. Que ce soit des indices relevés dans les cas d'AA-AR ou dans les cas d'AA-DR, nous remarquons que les données sont insuffisantes pour les considérer comme des marqueurs : l'approche onomasiologique ne peut se suffire à elle-même, seule une analyse sémasiologique pourra rendre compte des éléments favorisant l'interprétation d'une relation de discours plutôt que d'une autre.

9.2.4 Bilan

Nous proposons ci-dessous un tableau récapitulatif des différents types d'indices observés dans des configurations d'*Élaboration* dans notre corpus. Cette liste d'indices n'est certainement pas exhaustive. Elle pourra être complétée par l'examen des autres 1082 configurations des relations d'élaboration lorsque les annotations de référence pour les 42 textes du corpus seront disponibles.

Type d'indices	Descriptions/Exemples
Indices de type discursif	
Titres	titre référentiel, titre thématique...
Adverbiaux détachés	cadre de discours spatio-temporel, cadre thématique...
Amorces	amorce introduisant un nom d'éventualité avec un SN prospectif annonçant un découpage «en deux phases», «en deux temps» et/ou une manière «de la manière suivante», «de la façon suivante»...
Indices de type syntaxique	
SN	SN pluriel, SN démonstratif
Adjoints	gérondif, participe présent
Relatives	relative introduite par les relatifs <i>qui</i> , <i>que</i> , <i>dont</i> , <i>où</i>
Indices de type sémantique	
	<i>notamment, en particulier, entre autres, plus particulièrement, plus précisément, plus exactement, essentiellement</i>
Marqueurs structurels	<i>d'abord, puis, ensuite, ou bien... ou bien</i>
Adjectifs et adverbes numéraux	<i>premier, second, troisième, quatrième, un, deux, trois, quatre, premièrement, deuxièmement, troisièmement, quatrièmement...</i>
Corrélatifs anaphoriques	<i>le premier... le second, l'un... l'autre</i>
Adverbes de reformulation, d'inclusion, d'exemplification	<i>c'est-à-dire, y compris, par exemple</i>
Autres adverbes	<i>ainsi</i>
Marques ponctuationnelles	deux points, tirets
Prépositions	<i>avec, comme</i>

Une analyse sémasiologique de chacun de ces indices permettra de mettre au jour leur rôle dans la mise en œuvre de la relation d'*Élaboration*. Ces analyses pourront par la suite constituer un point de départ pour l'annotation et l'identification automatique de la relation.

9.3 Conclusion

Cette exploration des données ANNODIS nous a apporté des nouvelles données et des pistes pour l'étude de la relation d'*Élaboration*. Nous avons noté la complémentarité des approches sémasiologique et onomasiologique. Même si en théorie, un large corpus annoté en relations de discours pourrait suffire à mettre en corrélation indices et relations de discours, par exemple par le biais d'un apprentissage automatique, la pratique montre que les données ANNODIS sont inexploitable pour ce type d'expérience et que la meilleure façon de mettre en corrélation indices et relations de discours demeure l'approche sémasiologique.

Nous illustrons la complémentarité des deux approches dans la figure 9.18.

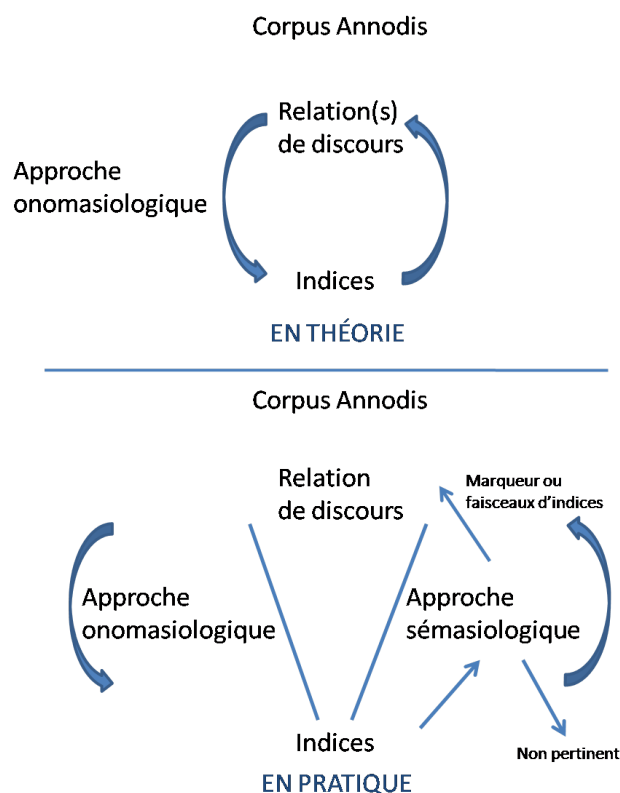


FIG. 9.18 – Complémentarité des approches onomasiologique et sémasiologique

Cette exploration nous a amenée à porter une réflexion sur la tâche d’annotation de relations de discours en corpus, sur les besoins et objectifs du linguiste en discours et sur la façon de construire ce corpus pour atteindre ces objectifs.

Les objectifs que nous nous sommes donnés au début de ce chapitre étaient doubles : d’une part, nous souhaitions valider nos indices mis au jour dans les chapitres de la partie 2 et d’autre part, nous souhaitions découvrir de nouvelles structures et de nouveaux indices en partant des annotations de la relation d’*Élaboration*.

Valider nos marqueurs Cet objectif n’a pu être atteint de façon satisfaisante pour plusieurs raisons.

La tâche d’annotation d’un corpus avec des relations de discours est une tâche lourde et complexe. Peu de données peuvent être annotées. Si l’on s’intéresse au fonctionnement d’un marqueur ou même d’une configuration de marqueur, il est donc nécessaire de focaliser notre attention et les annotations sur ces seuls marqueurs dans une perspective sémasiologique.

De plus, si l’on souhaite valider des analyses des fonctionnements discursifs des marqueurs, il est évidemment nécessaire de ne donner aucune indication aux annotateurs sur

ce qui est attendu. Il est donc proscrit, pour atteindre cet objectif, de fournir une liste de marqueurs, même potentiels, des relations de discours attendues.

Découvrir de nouvelles structures et de nouveaux indices Cet objectif n’a pu être atteint que partiellement.

Les accords entre annotateurs, très faibles, sur la relation d’*Élaboration*, mais également sur les autres relations annotées, ne permettent pas d’identifier des indices fiables. Les divergences d’annotation, si elles pointent quelquefois des lacunes dans les définitions du manuel, sont généralement le résultat d’un manque de formation et d’entraînement. L’expérience de la RST (Mann et Thompson, 2001) a montré qu’une annotation en relations de discours, aussi intuitive qu’elle se revendique, nécessite une longue période de formation et d’entraînement, de préférence par un expert en analyse du discours. Cette phase de formation a cruellement manqué à nos annotateurs pour lesquels la tâche d’annotation s’est avérée complexe³. Toute la difficulté réside, de notre point de vue, à trouver un juste milieu entre former et formater.

Ce dernier point nous amène à un autre objectif du projet ANNODIS qui est de faire un apprentissage automatique à partir des données recueillies. Il nous semble que les données obtenues ne permettront pas de mener à bien cet objectif. Dans ce cas, il faut partir du principe que le modèle employé est suffisamment bien détaillé pour permettre aux annotateurs d’accéder à la structure du discours. Nous ne nous situons plus dans un cadre de validation d’indices ou d’hypothèses mais nous les tenons pour acquis et amenons nos annotateurs à fournir l’annotation attendue pour mener à bien l’apprentissage automatique. Pour cet objectif, les données ici recueillies dans le projet ANNODIS nous permettront d’augmenter et d’améliorer les manuels. Une plus longue période de formation et d’entraînement sera alors nécessaire pour nos annotateurs.

Il semble qu’idéalement, il faudrait faire profiter aux annotateurs des analyses détaillées des indices et marqueurs des relations de discours. Il faut, de notre point de vue, fournir des descriptifs détaillés des fonctionnements des marqueurs pour guider les annotateurs plutôt que de leur donner une liste de marqueurs potentiels pour chaque relation.

³De plus, nous n’avons pas présenté les difficultés que les annotateurs ont rencontrées avec l’interface d’annotation dont l’ergonomie n’est pas adaptée aux besoins de cette annotation.

Chapitre 10

Identification automatique d'*Élaboration* en corpus

Sommaire

10.1 De la cohésion lexicale à la cohérence discursive	304
10.1.1 Répétition lexicale et chaînes lexicales	304
10.1.2 Indices lexicaux pour l'identification d' <i>Élaboration</i>	305
10.1.3 Choix de la ressource lexicale	307
10.2 Une expérience de combinaison	309
10.2.1 Combinaison d'indices	309
10.2.2 Principes de l'expérience et validation expérimentale	310
10.2.3 Bilan de l'expérimentation	312
10.3 Conclusions et perspectives	313

Dans ce dernier chapitre de la thèse, toujours dans une perspective de linguistique de corpus, nous souhaitons aborder la question de l'identification automatique de la relation d'*Élaboration*. La tâche d'identification automatique de cette relation est particulièrement difficile. D'une part, il existe peu de marqueurs de la relation et d'autre part ces marqueurs ont des fréquences faibles dans les textes. De plus, nous avons montré dans les chapitres précédents (à travers une approche sémasiologique et une approche onomasiologique) que l'interprétation de la relation fait généralement intervenir des sources d'informations diverses. Des faisceaux d'indices peuvent constituer des marqueurs de la relation d'*Élaboration*. Par ailleurs ces faisceaux font généralement intervenir des connaissances du monde et des informations liées à la sémantique lexicale. Ces informations sont particulièrement difficiles à repérer dans les textes, en l'absence de ressources lexicale et ontologique adéquates.

Proposer des solutions pour l'identification automatique de la relation permet de répondre à des problématiques pratiques, liées aux besoins de certaines applications en TAL. Cela permet également de contribuer à des problématiques théoriques, en étudiant les indices contribuant à l'interprétation d'une relation de discours et en collectant de nouveaux exemples, le tout en employant une méthode peu exploitée à notre connaissance.

La méthode, que nous présentons dans ce chapitre, fait intervenir la combinaison d'indices ambigus de la relation d'*Élaboration* à des indices de cohésion lexicale. Plus précisé-

ment, nous présentons les résultats obtenus lors d'une première expérience de combinaison de ces indices de cohésion lexicale à la construction gérondiva. Nous montrons que cette combinaison améliore la précision pour la tâche d'identification de la relation d'Élaboration.

Dans la section 10.1, nous présentons les indices de cohésion lexicale qui peuvent nous aider dans la tâche d'identification automatique de la relation d'Élaboration. Puis, dans la section 10.2, nous présentons notre expérience de combinaison des indices de cohésion lexicale à la construction gérondiva pour l'identification de la relation d'Élaboration. Notre hypothèse principale repose sur l'idée que les verbes et objets de la proposition principale et du syntagme gérondiva seront généralement repérables par des indices de cohésion lexicale dans les cas d'Élaboration et pas dans les autres cas.

10.1 De la cohésion lexicale à la cohérence discursive

Notre objectif est d'approcher une identification plus précise des élaborations dans les textes en faisant intervenir des indices de cohésion lexicale. Nous avons, à plusieurs reprises, mis en évidence le rôle des relations lexicales entre lexèmes et plus globalement entre types d'éventualités dans la mise en œuvre de la relation d'Élaboration, informations encodées par le prédicat *Subtyped_D* en SDRT. Ces informations sont particulièrement difficiles à repérer dans les textes, en l'absence de ressources lexicale et ontologique adéquates.

10.1.1 Répétition lexicale et chaînes lexicales

Harabagiu (1999) et Marcu (2000) ont proposé de tirer parti d'indices de cohésion lexicale pour l'identification automatique de la structure rhétorique des textes.

(Marcu, 2000) emploie une méthode qu'il qualifie lui-même de *coarse* (grossière). En l'absence de marques d'une autre relation de discours, la relation d'Élaboration ou la relation d'*Arrière-plan* est établie entre deux segments si ces derniers parlent de la même chose, i.e. s'ils partagent des mots communs. Dans le cas contraire, la relation de *Jointure* est établie. Le calcul proposé par Marcu repose sur la simple répétition lexicale. Après évaluation de son modèle, il conclut que cette méthode n'est pas suffisante.

Harabagiu (1999), quant à lui, emploie une méthode faisant intervenir une ressource lexicale, *WordNet*. *WordNet* est une base de données lexicales, développée par des linguistes de l'université de Princeton, pour l'anglais. Cette ressource contient des informations lexicales pour un ensemble de *synset* (groupe de mots dénotant un sens ou un usage particulier) de noms, de verbes, d'adjectifs et d'adverbes. Les sens des mots (et donc par extension les mots) sont liés entre eux par 13 relations lexico-sémantiques (par exemple l'hypéronymie, la synonymie, l'antonymie). Deux relations (la relation «implique» et la relation «est la cause de») organisent spécialement les verbes.

À partir de cette ressource, Harabagiu propose un algorithme qui dérive des chaînes lexicales rendant compte de la structure cohésive du texte. Nous illustrons une chaîne cohésive avec l'exemple (1) :

- (1) [Tuesday night's assassination of leaders of Swapo is a clear sign that the shooting is not over.]_a
 [Gunned down, Anton Lubowsky was the highest-ranking white person in Swapo.]_b
 [Les assassinats des leaders de Swapo jeudi dernier sont un signe que les fusillades

ne sont pas terminées.]_a
*[Anton Lubowsky, une des victimes, était un des blancs les mieux placés de Swapo.]*_b

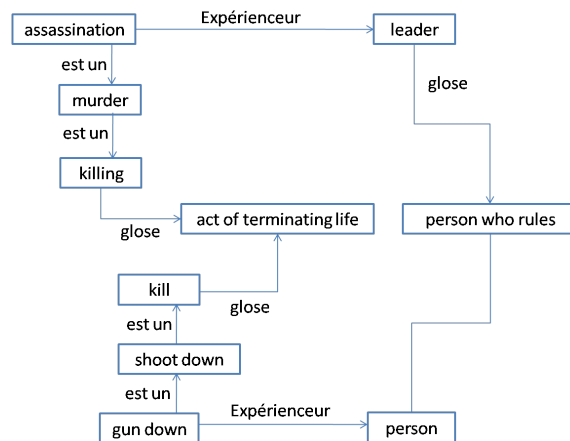


FIG. 10.1 – Un exemple de chaîne lexicale projetée sur (1)

Comme illustré en Figure 10.1, les mots des phrases (a) et (b) «assassination», «leader», «gun down» et «person» (dérivé de «Anton Lubowsky») sont liés par une chaîne lexicale faisant intervenir les relations d'hypéronymie, de glose et d'expérienceur.

Ces chaînes lexicales sont ensuite associées à des marqueurs, tels que *in others words*, pour la relation d'Élaboration, de sorte à repérer dans les textes des patrons de ces chaînes lexicales pour l'indentification d'une relation de discours donnée.

Pour l'identification d'Élaboration, les contraintes posées par Harabagiu pour son identification sont simples : il faut qu'il existe des chaînes lexicales entre certaines têtes de syntagme dans les deux phrases. Pour identifier un *Contraste*, la chaîne lexicale doit impliquer la relation d'antonymie. Nous voyons que les chaînes lexicales permettant d'identifier la relation d'Élaboration sont, ici également, fortement sous-spécifiées. L'exemple (1) d'Harabagiu illustre toutefois l'importance de la relation d'hypéronymie («est un») pour l'identification de la relation d'Élaboration.

WordNet semble être une option intéressante pour notre tâche d'identification de la relation d'Élaboration. Pour l'anglais, cette ressource présente une couverture assez large de relations typées. En revanche, elle ne met pas en évidence les relations lexicales non classiques (alors que ces dernières sont pertinentes pour l'identification de l'Élaboration comme nous le verrons dans la section suivantes) et n'est pas disponible pour le français.

10.1.2 Indices lexicaux pour l'identification d'Élaboration

Dans le cadre de la SDRT, le lexique est une source d'information importante (mais pas exclusive) pour l'inférence du prédicat *Subtype_D* (Cf. section 3.2.1 du chapitre 3 page 85 pour la présentation du prédicat).

À première vue, une ressource fournissant des hypéronymes paraît être la ressource appropriée pour détecter automatiquement des élaborations. Mais, comme nous l'avons

vu tout au long de cette thèse, les types de relations se manifestant dans les structures élaboratives sont plus variés que la seule hypéronymie. De plus, étant donné que ces relations émergent en discours, les phénomènes lexicaux à l'œuvre ne sont pas ceux que l'on trouve dans les ressources lexicales classiques. Ces relations peuvent s'établir en discours en étant fermement reliées à une énonciation spécifique (Mortureux, 1993). Nous illustrons la variété des phénomènes lexicaux à l'œuvre avec l'exemple (2).

- (2) [Un **véhicule** a effectué une spectaculaire **sortie** de **route**, hier vers 18 h 15, sur l'A36.]₁ [La **voiture** circulait dans le sens Mulhouse-Montbéliard]₂ [lorsqu'après être passée à hauteur du 35e RI,]₃ [elle a **quitté** la **chaussée** sur sa droite.]₄ (L'Est républicain)

L'interprétation d'*Élaboration* (1,4) fait intervenir un ensemble de relations lexicales présentes entre les segments <1> et <4>. Par exemple, la reconnaissance de l'identité des sujets (ou expérienceurs pour leur rôle thématique) entre les deux segments passe par la reconnaissance d'une relation d'hypéronymie «véhicule»/«voiture» (antécédent du pronom «elle» du segment 4). Les syntagmes verbaux des deux segments font également intervenir des relations lexicales : la relation de méronymie entre «route» et «chaussée» et une relation plus difficile à caractériser entre «sortie» (de «effectuer une sortie») et «quitter». Le nom «sortie» dénote, en quelque sorte, l'action décrite par le verbe. La dernière relation, établie entre un nom et un verbe, est un type de relation intercatégorielle.

Si les deux premières relations d'hypéronymie et de méronymie peuvent être accessibles *via* des ressources lexicales. La troisième, en tant que relation intercatégorielle, n'est généralement pas accessible *via* une ressource lexicale classique.

L'exemple (3) illustre également ce type de relation intercatégorielle.

- (3) (...) [qui rappelle la **vocation** des bénévoles de l'association :]₃₂ [être un **soutien** pour la paroisse,]₃₃ [apporter une petite **contribution** financière aux travaux grâce aux manifestations et aux dons,]₃₄ [**accomplir** de multiples **tâches** et démarches touchant aux bâtiments paroissiaux,]₃₅ [**contribuer** à la convivialité entre les paroissiens.]₃₆ (L'Est républicain)

En (3), les types d'éventualités décrites dans les segments <33>, <34>, <35> et <36> sont des sous-types du type de l'éventualité décrite en <32> : «vocation des bénévoles de l'association». L'interprétation de la relation d'*Élaboration* (*Élaboration*(32,[33-36])) fait intervenir des liens lexicaux s'établissant entre «vocation» et les mots suivants : «soutien», «accomplir», «tâche», «contribuer»... Ces liens établis en discours (et basés sur des connaissances du monde en plus de connaissances lexicales) n'apparaîtront probablement pas dans une ressource générique, tout simplement parce que ces liens intercatégoriels ne sont généralement pas recensés dans les ressources lexicales classiques.

Ces observations orientent le choix de la ressource lexicale que nous souhaitons utiliser pour notre tâche d'identification automatique de la relation d'*Élaboration*.

10.1.3 Choix de la ressource lexicale

Les liens sur lesquels se base l'inférence de la relation d'*Élaboration* sont plus souples que la relation de sous-type lexical. Dans le but de détecter automatiquement ces liens, pour améliorer la détection des élaborations, le choix de la ressource est crucial : elle doit contenir, au maximum, les liens, présentés ci-dessus, qui n'apparaissent pas dans les ressources classiques.

Nous avons porté notre attention sur une ressource bâtie en corpus, qui prend en compte une variété de relations de proximité sémantique entre mots d'une même catégorie morpho-syntaxique ou de catégories morpho-syntaxiques différentes. Cette ressource, les *Voisins de Wikipedia*, est construite au moyen d'une analyse distributionnelle. Le principe de cette analyse est de rapprocher des paires de mots selon les contextes qu'ils partagent, suivant les hypothèses de Harris (1968). Ces paires de mot partagent des affinités de second-ordre, c'est-à-dire que les mots de la paire n'apparaissent pas ensemble dans le corpus (contrairement aux collocations) mais que leurs environnements d'apparition sont similaires (Grefenstette, 1994). Les relations lexicales ainsi mises en évidence sont des relations paradigmatiques (*Cf.* Tableau 10.1) (contrairement aux collocations qui mettent en évidence des relations syntagmatiques en rapprochant des mots qui sont souvent associés dans les textes) (*Cf.* exemples (4), (5) et (6)).

- (4) voix suave
- (5) courir vite
- (6) entraîner des conséquences

Les *Voisins de Wikipedia* ont été générés par le programme *Syntex-Upéry* développé par Bourigault (2002) à partir d'une importante collection d'articles de *Wikipedia* contenant 194 millions de mots.

La première étape de la chaîne consiste en une analyse syntaxique. Les relations syntaxiques sont représentées sous forme de triplet <gouverneur, relation, dépendant>. Pour la phrase «Clémentine circule à bord d'une voiture», les triplets suivant sont repérés : <circuler, sujet, Clémentine> et <circuler, à bord de, voiture>. Puis, les triplets ainsi obtenus sont transformés en couple <prédicat, argument>, soit respectivement <circuler_suj, Clémentine> et <circuler_à bord de, voiture>. À partir de ces paires, il est possible de rapprocher tous les arguments partageant les mêmes prédicats comme *véhicule* et *voiture* et tous les prédicats partageant les mêmes arguments comme *sortie* et *quitter* :

Par exemple, les voisins *véhicule* et *voiture* sont rapprochés sur la base des contextes partagés suivants :

- circuler_à bord de
- pare-brise_de
- portière_de
- capot_de
- percuter_objet
- immatriculation_de
- conduire_objet

Les voisins *sortie* et *quitter* sont rapprochés sur la base des contextes partagés suivants :

Synonymie	Antonymie	Hypo/hyperonymie	Co-hyponymie	Autre relation
blanc/pâle	clair/obscur	voiture/véhicule	voiture/train	peintre/tableau
agressif/violent	matériel/spirituel	chien/animal	lundi/mardi	ordinateur/internet
facile/simple	stable/instable	marcher/se déplacer	chat/cheval	université/enseignement

TAB. 10.1 – Typologie de voisins distributionnels

- territoire
- système éducatif
- conservatoire
- hôpital
- autoroute

La ressource ainsi obtenue comporte 4 millions de paires, couvrant un large éventail de relations, comme illustré dans le tableau 10.1.

La figure 10.2 illustre la projection de ces voisins sur un extrait de texte.

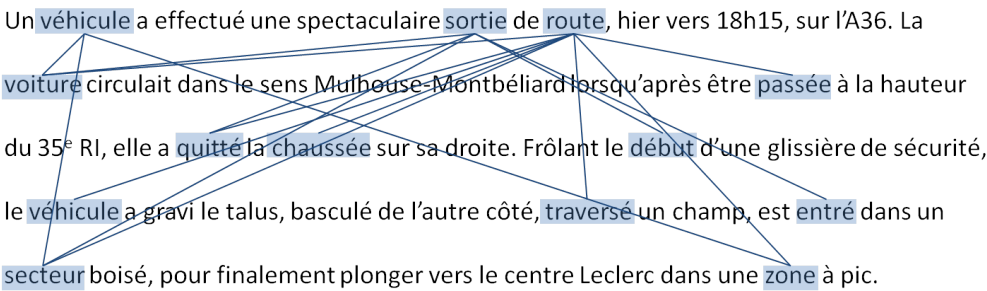


FIG. 10.2 – Un exemple des liens de voisinage projetés sur (2)

Les liens, que nous avons précédemment signalés comme aidant à l'inférence de la relation d'Élaboration en (2), sont identifiés : *véhicule/voiture*, *sortir/quitter* et *route/chaussée*. D'autres liens participant à la cohésion globale sont également observés mais ces liens ne sont pas directement impliqués dans l'inférence d'Élaboration : *route/véhicule*. Enfin certains liens ne sont pas pertinents dans ce contexte discursif et ne participent pas directement à la cohésion globale du texte : *route/traverser* (puisque'il ne s'agit pas de traverser la route, mais de traverser un champ).

La nature pléthorique (Cf. tableau 10.1 illustrant les relations sémantiques observables dans les paires des voisins distributionnels) de cette ressource rend difficile son exploitation sans avoir un contrôle sur les liens projetés. Bien que les relations qui sont pertinentes pour la détection des élaborations soient mises en évidence par la projection des *Voisins de Wikipedia*, bien d'autres liens de voisinage interfèrent, rendant impossible l'interprétation directe des liens de voisinage pertinents pour notre identification de la relation

d'*Élaboration*. Il nous apparaît alors nécessaire de définir des marqueurs plus restrictifs, en prenant en compte d'autres critères que la simple présence de liens de voisinage. Notre idée est d'utiliser notre liste d'indices pour cibler des zones où les liens de voisinage pourraient discriminer des structures élaboratives des autres structures.

Dans la section suivante, nous allons illustrer ce type de combinaison en présentant l'expérimentation que nous avons menée en collaboration avec Clémentine Adam (Adam et Vergez-Couret, 2010). Nous avons proposé de combiner la construction gérondive et des indices de cohésion lexicale obtenus avec les *Voisins distributionnels* pour identifier avec plus de précision des constructions gérondives dans lesquelles s'établit une relation d'*Élaboration*.

10.2 Une expérience de combinaison

10.2.1 Combinaison d'indices

L'idée de combinaison peut être abordée de deux façons : d'un côté nous pensons que la combinaison de nos indices ambigus de la relation à des indices de cohésion lexicale va fournir un patron d'identification plus fort et plus fiable ; de l'autre, la combinaison s'impose aussi par la nature de la ressource des *Voisins de Wikipedia*, qui recense une large variété de relations de similarité lexicale.

Pour mener cette première expérimentation, le gérondif est apparu comme le candidat idéal. Certains syntagmes gérondifs sont des élaborations de la proposition principale mais pas tous. L'identification des segments à relier est déterminée par la syntaxe : le segment à attacher est le syntagme gérondif et le segment cible la proposition principale. Enfin, dans le cadre d'une identification automatique, le gérondif peut être extrait de façon fiable avec l'analyseur syntaxique, SYNTAX.

Pour rappel, la construction gérondive établit un lien de subordination syntaxique entre deux verbes : deux procès sont ainsi liés. Les différentes valeurs sémantiques exprimées par le syntagme gérondif ne sont pas véhiculées par le gérondif lui-même mais dépendent de la combinaison des deux verbes liés. L'interprétation se fait *a posteriori* et elle est déterminée par les relations sémantiques en jeu entre les verbes et d'autres éléments donnés par le contexte (Halmøy, 1982).

Nos analyses du chapitre 7 ont révélé que le syntagme gérondif peut être attaché à la proposition principale par plusieurs relations de discours : *Élaboration*, *Résultat*, *Explication*, *Conditionnel* et *Circonstance accompagnante* (Cf. Conclusion du chapitre 7).

L'hypothèse principale de cette expérience repose sur l'idée que les verbes et objets de la proposition principale et du syntagme gérondif seront généralement des voisins dans les cas d'*Élaboration* (verbes et objets mettant en jeu des relation d'hypéronymie ou de manière) et pas dans les autres cas (verbes et objets ne mettant en jeu aucune relation). Pour cette expérimentation, nous adoptons une vision étroite de la relation d'*Élaboration*. Autrement dit, nous ne considérons pas la relation de *Circonstance Accompagnante*, que nous souhaitons justement écarter. Nous cherchons les élaborations de manière d'un procès en cours de déroulement, les élaborations mettant en relation une réalisation concrète avec

la description d'un procès abstrait et/ou métaphorique et les élaborations mettant en relation un verbe de dire ou de déplacement avec un hyponyme de ces derniers.

Nous illustrons notre hypothèse avec les deux exemples (7) et (8), pour lesquels nous mettons en relation une analyse SDRT et l'absence ou présence de voisins distributionnels.

- (7) United Fruit Company investit dans le pays, en achetant des parts dans le chemin de fer, l'électricité et le télégraphe. (Wikipedia)
- (8) Dans la ville de Koriko, Kiki, accompagnée de son chat noir Jiji, va distribuer des colis en volant sur son balai, grâce à ses pouvoirs. (Wikipedia)

En (7), la proposition principale introduit l'événement «investir» et le syntagme gérondif introduit l'événement «acheter des parts». Le type de l'événement du syntagme gérondif est un sous-type de celui de la proposition principale. La relation d'*Élaboration* est donc inférée. Du point de vue des voisins, cette relation entre «investir» et «acheter des parts» pourrait être identifiée *via* la paire de voisins *investir/acheter*.

En (8), la proposition principale et le syntagme gérondif décrivent deux procès se déroulant simultanément. Le gérondif peut être remplacé par la forme «tout en volant sur son balai». La relation de *Circonstance accompagnante* est ici inférée. Du point de vue des voisins, notre hypothèse est que les verbes et objets de la proposition principale et du syntagme gérondif ne seront pas des voisins du fait de la relation fortuite s'établissant entre les deux procès.

10.2.2 Principes de l'expérience et validation expérimentale

Pour cette expérience, notre but est d'obtenir la meilleure précision dans l'identification de la relation d'*Élaboration*. Nous commençons par exclure les syntagmes gérondifs introduits par *tout* qui expriment typiquement des circonstances accompagnantes de la proposition principale.

Nous annotons donc les verbes et objets voisins dans les constructions gérondives selon deux stratégies :

- La proposition principale et le syntagme gérondif ont deux verbes voisins, Cf. (9) (*contribuer/envoyer*) (ces candidats sont notés VV (verbes voisins)).
 - La proposition principale et le syntagme gérondif ont deux verbes et deux objets voisins, Cf. (10) (*élargir/englober* et *empire/monde*) (ces candidats sont notés VVOV (verbes voisins et objets voisins)).
- (9) ... et les villages **contribuaient** également à ce grand projet religieux **en envoyant** des vivres. (Wikipedia)
 - (10) Les Skrulls (...) **élargissent** leur *empire* **en englobant** dans celui-ci les *mondes* moins avancés qu'ils rencontrent. (Wikipedia)

Pour cette expérience, nous avons annoté automatiquement 1/5 du corpus *Wikipedia* avec les *Voisins de Wikipedia*

Tous les paires <proposition principale, syntagme gérondif> sont extraites et trois sous-ensemble sont formés : l'ensemble GSV (construction gérondive sans lien de voisinage) et, selon les deux stratégies présentées ci-dessus, les deux sous-ensembles VV et VVOV sont

extraits, en annotant les verbes voisins (VV) et les verbes et objets voisins (VVOV). Ainsi, nous avons la répartition suivante :

GSV	VV	VVOV
18571	375	193

Le nombre de candidats pour l'identification de la relation d'*Élaboration* est faible au vu de la taille initiale du corpus. Cependant, dans l'état actuel des recherches sur cette relation, définir un faisceau d'indices fiable constitue une amélioration significative, même si le nombre d'occurrences est faible.

Chaque construction gérondive a été indépendamment annotée par Clémentine Adam et nous-même en relations de discours. 314 constructions gérondives ont été annotées, approximativement 100 pour chaque cas (GSV, VV, VVOV). Chaque extrait nous a été présenté de façon aléatoire avec pour question : *Est-ce que le syntagme gérondif est une élaboration de la proposition principale ?*

L'accord entre annotateurs est de 89% (280 cas d'accord et 34 cas de désaccord). Le coefficient du Kappa κ est de 0,70. L'accord entre deux annotateurs résultant de la somme d'une composante «aléatoire» et d'une composante d'accord «véritable», le Kappa permet de mesurer la qualité de l'accord réel entre les deux annotateurs, i.e. de s'assurer que l'accord entre les deux annotateurs n'est pas dû au simple hasard. Le Kappa pour cette annotation montre un accord de *modéré à bon*. Il donne un aperçu à la fois la difficulté de la tâche tout en autorisant une automatisation de celle-ci.

Nous sommes revenues sur les 34 exemples pour lesquels nous avons été en désaccord afin de proposer une annotation de référence la plus fiable possible et d'analyser les types de variation entre les deux annotations. La discussion nous a permis de trouver un accord sur la majorité des cas. Au final, notre avis est resté partagé sur 9 cas, notamment des cas où deux interprétations sont possibles. Ces 9 cas marginaux ont été retirés de l'annotation de référence.

Le tableau 10.2 résume les résultats obtenus pour toutes les constructions gérondives annotées.

	Nombre Total	Nombre Annoté	Elab.	Pas Elab.	Précision	Intervalle de confiance
GSV	18571	102	62	40	60.8%	9.45%
VV	375	100	81	19	81.0%	6.59%
VVOV	193	104	99	5	95.2%	2.8%

TAB. 10.2 – Résultats

Ces résultats confirment que la construction gérondive est un marqueur ambigu étant donné que 60% des cas sont des élaborations. Le nombre d'extraits annotés est faible comparé à tous les gérondifs extraits, d'où l'intervalle de confiance assez élevé. Celui-ci indique une marge d'erreur s'élevant à presque 10%. Toutefois, les différences obtenues entre

le gérondif sans voisins et nos deux stratégies sont suffisamment importantes pour assurer que nos deux stratégies permettent d'identifier avec plus de précision des élaborations. Avec la première stratégie (vv), 81 % des cas sont des élaborations. La seconde stratégie est encore plus fiable avec 95% de précision. Ces résultats, illustrés par les exemples de (11) à (14), sont prometteurs pour envisager une identification plus précise de la relation d'*Élaboration*.

- (11) Elle **décide** de se venger **en prenant** un faux amant en la personne de Philippe.
- (12) Le 4 février 1794, la Convention **ratifia** cette décision dans l'enthousiasme **en abolissant** l'esclavage dans toutes les colonies françaises.
- (13) La musique de La Talvera **crée** des *liens* dans la contemporanéité **en jetant** des *ponts*, d'un continent à l'autre, d'un temps à un autre.
- (14) Selon la Chronique de Nantes, Hoël «**pilla** le *territoire* de Rennes **en brûlant** les *maisons* jusque près de la cité».

Les cas où notre combinaison d'indices a échoué ont également été analysés. Dans peu de cas, l'échec résulte d'un lien de voisinage non pertinent, comme illustré en (15).

- (15) Ils *marchent* la campagne en *incendiant* toutes les habitations. (Wikipedia)

Les verbes *marcher* et *incendier* sont voisins mais se révèlent non pertinents dans ce contexte.

Dans certains cas, un marqueur d'une relation de discours différente d'*Élaboration* est observée, comme illustré en (16).

- (16) Le roi d'Espagne lui accorda une décoration qu'il accepta, *mais* en refusant la pension qui y était attachée. (Wikipedia)

La présence de certains marqueurs, comme le marqueur *tout + gérondif*, indique une autre relation de discours que l'*Élaboration*. Nous pourrions améliorer notre détection en filtrant tous les syntagmes gérondifs introduits par un marqueur d'une autre relation de discours, comme le marqueur de contraste *mais*.

10.2.3 Bilan de l'expérimentation

Cette expérience permet de montrer qu'une combinaison d'indices, la construction gérondive et les voisins distributionnels, permet d'améliorer l'identification de la relation d'*Élaboration*.

L'intérêt des indices de cohésion lexicale pour l'organisation structurelle du discours est communément accepté mais ces indices sont souvent négligés dans les applications de TAL du fait qu'il est difficile de les capturer dans les textes. Cette expérience montre la validité des *Voisins de Wikipedia* pour cette tâche d'identification de la relation d'*Élaboration*.

Il est vrai que notre approche ne permet de détecter, avec une bonne précision, que quelques cas d'*Élaboration* dans l'ensemble du corpus (environ 550 cas). Mais notre approche permet, à partir d'un indice ambigu, de construire un nouvel indice, ou plus exactement un faisceau d'indices, améliorant la détection de l'identification de l'*Élaboration*.

Nous souhaitons, pour le futur, évaluer notre méthode d'identification de la relation en combinant d'autres indices mis au jour dans cette thèse avec les voisins distributionnels. Des adaptations de notre méthodologie devront être proposées selon les types d'indices considérés.

Les *Voisins de Wikipedia* sont un exemple type de ressource endogène pour laquelle le corpus est à la fois l'objet du traitement et la source des informations acquises. La question de la portabilité de la ressource se pose donc inévitablement. Pour cette expérience, nous avons employé la ressource sur le corpus qui a servi à sa construction. Pour notre tâche, les relations lexicales recherchées sont celles-là mêmes qui ont été mises en place dans le discours, il nous paraît donc plus prudent d'utiliser la ressource des voisins distributionnels obtenue sur le corpus dans lequel nous cherchons à identifier la relation d'*Élaboration*. Si cette contrainte apparaît comme moins importante pour des corpus de «langue courante» comme *Wikipedia* ou *L'Est Républicain*, elle est cruciale dès qu'on s'intéresse à des corpus de «langue de spécialité» pour lesquels on peut s'attendre à la mise en œuvre de relations lexicales spécifiques au domaine considéré.

10.3 Conclusions et perspectives

Les expériences présentées dans ce chapitre apportent des données empiriques pour confirmer que les indices de cohésion lexicale représentent une source pertinente pour la reconnaissance de la relation d'*Élaboration*. Cela confirme, empiriquement, l'intuition que la cohérence discursive se construit, en partie, à partir de la structure cohésive du texte.

Notre expérience, dont les résultats sont encourageants, nous motive à poursuivre des expériences de combinaison d'indices ambigus et des voisins distributionnels pour contribuer à augmenter la précision de l'identification de la relation d'*Élaboration*.

Nous pensons avoir validé les *Voisins de Wikipedia* comme des indices pertinents pour identifier la relation d'*Élaboration*. Nous pensons donc à mettre en place une expérience à plus grande échelle.

1. Pour un indice ambigu X , nous proposons de les annoter automatiquement dans un corpus.
2. Puis nous proposons une annotation manuelle (i) des segments liés par l'indice ; (ii) de la relation établie entre les deux segments : *Élaboration* ou non.
3. Nous proposons ensuite d'analyser syntaxiquement le corpus (à l'aide du programme *Syntex*) et de projeter tous les liens de voisinage existant entre les mots des deux segments automatiquement.

4. Par le biais d'un apprentissage automatique, il est alors possible de mettre en corrélation l'annotation d'*Élaboration*, les liens de voisinage et la position syntaxique des mots voisins afin de découvrir des patrons pour l'identification automatique de la relation.

À terme, notre but serait de développer les faisceaux d'indices ainsi mis au jour pour envisager la reconnaissance de la relation d'*Élaboration* là où aucun marqueur lexical et/ou syntaxique n'est employé.

Cette proposition ajoute une méthode, empruntée à la linguistique de corpus, pour déterminer des patrons d'identification automatique. Cette proposition ne vient pas remplacer

la précédente, empruntée à la linguistique descriptive. Les deux sont complémentaires et se nourrissent l'une l'autre, Cf. Figure 10.3.

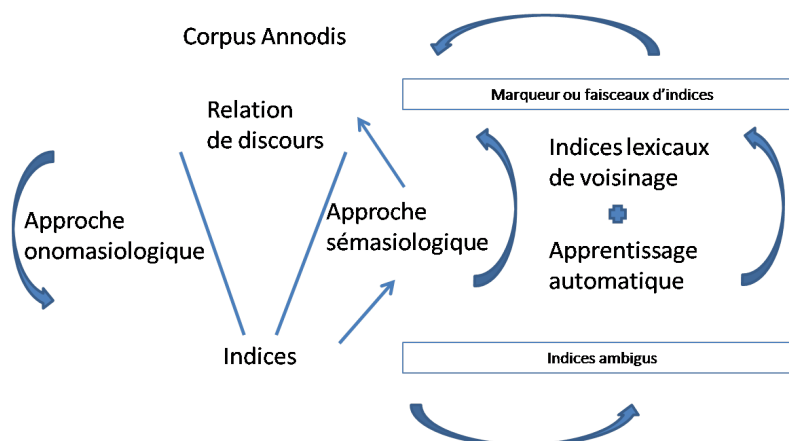


FIG. 10.3 – Complémentarité des approches onomasiologique et sémasiologique

Les résultats de l'approche sémasiologique sont nécessaires pour apporter des spécifications pour l'annotation des indices ambigus (ce que nous avons illustré avec les résultats de l'approche sémasiologique du chapitre 7 qui nous ont guidé pour notre annotation des élaborations). Et les résultats obtenus par la recherche de patrons d'identification donnent des validations empiriques venant enrichir les résultats de l'analyse sémasiologique (ainsi notre expérience a validé, empiriquement, que la relation d'*Élaboration* peut être marquée lexicalement).

Conclusion

Au terme de ce travail, nous pouvons récapituler nos avancées sur l'exploration de la relation d'*Élaboration* dans les quatre axes de travail considérés dans notre thèse : la description linguistique, la modélisation linguistique, l'exploration en corpus et les traitements informatiques. Nous avons proposé une méthodologie ayant pour objectif d'articuler ces quatre axes pour faire progresser nos connaissances sur la relation.

Pour approcher la relation d'*Élaboration*, nous avons d'abord adopté une démarche onomasiologique que nous avons qualifiée d'*introspective*. Il s'agissait, en partant de la relation, de déterminer quelles pouvaient être ses réalisations linguistiques en s'appuyant d'une part sur notre intuition et d'autre part sur l'exploration sporadique de textes (Partie 2). L'étude d'un marqueur potentiel nous a naturellement entraînée à en considérer d'autres. Par exemple, l'intérêt porté à un adverbe comme *notamment* nous a amenée à considérer d'autres adverbes paradigmatiques. Puis l'étude des adverbes paradigmatiques nous a conduite à l'analyse des comparatifs nus *plus particulièrement*, *plus précisément*. La démarche sémasiologique a, ensuite, permis de faire progresser les analyses des marqueurs potentiels considérés et d'enrichir un modèle du discours.

À la fin de nos quatre années de thèse, la mise à contribution du corpus ANNODIS annoté en relations de discours nous a naturellement entraînée à proposer une méthodologie plus systématique de recherche d'indices (Partie 3). Son exploration préliminaire a également permis de montrer que les analyses sémasiologiques sur les indices d'une relation de discours permettraient, si elles étaient prises en considération, d'améliorer les manuels d'annotation et, par conséquent, la qualité des annotations.

Pour les travaux de description du fonctionnement des marqueurs potentiels, dont une part a donné lieu à des formalisations, nous nous sommes placée dans le modèle de la SDRT. Ce cadre nous a permis d'étudier nos données en cherchant à déterminer, avec les outils du modèle, les éléments responsables de la mise en œuvre de la relation.

Nous avons fourni des descriptions détaillées sur de nombreux exemples attestés et fait progresser des analyses linguistiques sur les marqueurs lexicaux *notamment*, *en particulier*, *plus particulièrement*, *plus précisément* (chapitre 5), les structures énumératives (chapitre 6) et sur la construction gérondive (chapitre 7). Ces analyses ont eu pour but immédiat d'enrichir le modèle de la SDRT concernant la relation d'*Élaboration*. Nous avons également fourni dans cette thèse de nombreuses analyses SDRT de mini-discours attestés avec les représentations graphiques correspondantes.

Les règles d'inférences proposées dans cette thèse viennent directement enrichir le modèle de la SDRT.

Nous avons, d'une part, enrichi le modèle avec l'introduction de deux nouveaux prédicats encodant de nouvelles configurations permettant de déclencher *Élaboration* : *Subevent_D* et *Moyen_D*. Le prédicat *Subevent_D* encode des informations de type méronymique sur le possible découpage d'une éventualité en sous-éventualités (chapitre 3). Le prédicat *Moyen_D* encode des informations sur la façon dont on peut accomplir une action (chapitre 7).

Nous avons ensuite proposé des règles d'inférence basées sur la présence de plusieurs indices. Nous n'avons pas mis au jour un marqueur type de la relation de discours *Élaboration*. En revanche, des indices combinés, comme un adverbe et une position, forment des marqueurs de la relation d'*Élaboration*. Les adverbes *notamment*, *en particulier* et *plus précisément* en position initiale constituent des marqueurs de la relation. Concernant les structures énumératives, nous avons montré que les règles d'inférence se cumulent jusqu'à l'inférence de la relation d'*Élaboration*. D'abord, on reconnaît une structure énumérative par le biais de la reconnaissance d'une amorce et d'un marqueur de premier item. Une fois la structure énumérative mise en évidence, l'inférence de la relation d'*Élaboration* est établie en présence d'une amorce de type événementiel.

L'observation de ces marqueurs a, par ailleurs, mis en évidence des types d'élaboration, jusqu'alors non inférables, comme *plus particulièrement* qui peut signaler une relation d'*Élaboration* du type *du général au particulier*; *plus précisément* qui peut signaler une relation d'*Élaboration* de type *reformulation* et la construction gérondive qui peut signaler une relation d'*Élaboration* de type *circonstance accompagnante*. L'observation de *plus précisément* a également mis en évidence des cas de *méta-Élaboration*, notée *Élaboration**.

Nous avons proposé diverses règles d'inférence correspondant à des cas de figures observés dans nos données. Notre objectif n'est pas de multiplier les relations de discours. Fidèle à l'approche réductrice proposée par la SDRT, nous partageons le point de vue de Hovy et Maier (1991) selon lequel il peut être suffisant de considérer une seule relation d'*Élaboration* pour exprimer tous ces concepts. Des raffinements de la relation peuvent éventuellement être proposés selon les besoins et les buts d'une analyse donnée.

Pour systématiser la recherche d'indices de la relation, l'approche onomasiologique, présentée dans la partie 3, a fourni des données qui permettent, d'une part d'augmenter notre liste d'indices de la relation d'*Élaboration* et d'autre part d'envisager des améliorations pour des annotations futures en relations de discours.

Grâce aux données du corpus ANNODIS, nous avons pu aborder avec un nouveau point de vue les réalisations linguistiques de la relation. L'exploration de ces données (plus exactement d'une double annotation en relations de discours pour chaque texte) a permis également de faire surgir des configurations et des indices (par exemples les propositions relatives ou les titres) pour lesquels les annotations ont été divergentes (chapitre 9).

Le corpus ANNODIS permet, en quelque sorte, de se passer de la phase d'introspection que nous avons adoptée au début de notre thèse en l'absence d'une telle ressource. Il permet d'adopter une approche systématique (dans les limites du corpus) de recherche d'indices.

Par ailleurs, cette première expérience d'annotation a confirmé qu'il est, d'une part,

difficile d'obtenir des annotations en relations de discours fiables et, d'autre part, coûteux d'obtenir de telles annotations en grande quantité.

Un des résultats de notre travail a été de proposer une méthodologie pour l'exploration d'une relation de discours et l'exploitation du corpus. La première étape consiste à réunir des indices de la relation. Ce qui peut apparaître comme un point faible du corpus (annotation non fiable) est finalement ce dont nous avons tiré bénéfice en mettant au jour des indices et des configurations qui ont causé des difficultés aux annotateurs et qui nécessitent donc des analyses d'experts.

La deuxième étape consiste à adopter une approche sémasiologique pour l'analyse de ces indices dans le but d'enrichir le modèle (en faisant progresser des analyses linguistiques sur ces derniers) et d'améliorer des manuels pour une annotation future. Nous avons proposé que les futurs manuels ne doivent pas uniquement comporter des descriptions des relations de discours et des listes de marqueurs potentiels. Ils doivent de notre point de vue comporter des listes d'indices accompagnés de spécifications d'annotation. Les rédacteurs de ces manuels pourraient bénéficier pour la rédaction de ces spécifications des résultats de l'approche sémasiologique (établir les contextes et les contraintes pesant sur un indice dans la mise en œuvre d'une relation de discours). Pour autant, nous ne garantissons pas que les futures données seront fiables, mais espérons seulement qu'elles le seront davantage. La tâche d'annotation en relations de discours n'est pas triviale et de nombreux aller-retours entre approches onomasiologique et sémasiologique nous paraissent nécessaires.

Dans une perspective d'identification automatique ou semi-automatique de la relation, toutes nos analyses ont pointé vers l'idée qu'un marqueur ou un indice ne suffit généralement pas et qu'il est nécessaire de mettre au jour des configurations d'indices. Notre première expérience de traitements informatiques pour l'identification de la relation d'*Élaboration* repose d'une part sur l'idée que la combinaison de deux indices ambigus constitue un indice plus fort pour le repérage et d'autre part sur l'idée que des indices de cohésion lexicale peuvent désambiguïser nos indices lexicaux et syntaxiques ambigus (chapitre 10). Le premier patron que nous avons testé – combinaison de la construction gérondive et d'indices de cohésion lexicale – a fourni des résultats très prometteurs. Ce patron repose sur l'hypothèse que des indices de cohésion lexicale rentrent en jeu dans la mise en œuvre de la relation. Cette méthode, empruntée au TAL, a permis de tester à plus grande échelle cette hypothèse dans la mise en œuvre de la relation d'*Élaboration*. L'indice obtenu par cette combinaison a permis d'identifier avec plus de précision des constructions gérondives mettant en jeu la relation d'*Élaboration*, validant ainsi nos hypothèses de départ.

Nous sommes consciente de ne pas avoir épuisé l'étude des réalisations linguistiques de la relation d'*Élaboration*. De nombreuses perspectives s'inscrivent donc directement dans la lignée de nos travaux.

À court terme, nous pourrions commencer par analyser toutes les annotations du corpus pour la relation d'*Élaboration* et d'*Élaboration d'entité*, en appliquant la méthodologie que nous proposons, afin d'identifier d'autres indices des deux relations.

Tous les indices doivent alors faire l'objet d'une analyse sémasiologique discursive détaillée pour déterminer leur rôle dans la mise en œuvre des relations d'élaboration. Et cela dans le double objectif d'enrichir les modèles du discours et de contribuer à l'amélioration des manuels d'annotation pour une future campagne d'annotation.

Concernant ces indices, dans une perspective de linguistique de corpus et dans l'objectif du repérage automatique de la relation, il serait pertinent de déterminer, pour chacun d'eux, leur fiabilité dans la reconnaissance de la relation d'*Élaboration*, i.e. dans quel pourcentage chaque indice signale la relation. Cette étude pourrait également être menée sur des corpus de genres différents afin de déterminer si les fonctions discursives des marqueurs varient d'un genre à l'autre.

Pour poursuivre cette étude sur la relation d'*Élaboration*, en tirant bénéfice de nos analyses, nous imaginons la mise en place d'une annotation similaire à celle proposée par l'approche descendante du projet ANNODIS. Il s'agit de se concentrer sur l'annotation des relations d'élaboration en projetant sur le corpus les indices identifiés (marqueurs lexicaux, structures syntaxiques, indices de cohésion lexicale...), auxquels seraient apportées des instructions d'annotation. La densité de ces indices permettraient de pointer vers des structures élaboratives. Une fois les annotations validées, celles-ci pourraient nous permettre d'accéder à de nouveaux indices. Nous concentrer sur un seul groupe de relations de discours et des indices, préalablement étudiés, permettraient d'augmenter la quantité et la qualité des annotations de celles-ci. Nous pourrions, ainsi, envisager des méthodes d'apprentissage automatique pour mettre au jour des combinaisons d'indices pointant fiablement vers des structures élaboratives.

Cette méthode, que nous présentons pour la relation d'*Élaboration*, peut, nous semble-t-il, être transposée à toutes les relations de discours. Car même si nous avons présenté certaines de ces relations comme ayant des marqueurs prototypiques, toutes les réalisations de ces relations ne sont pas forcément signalées par ces derniers. En adoptant une démarche onomasiologique, que l'on peut systématiser grâce aux annotations du corpus ANNODIS, il est possible de repérer d'autres indices de ces relations. Le rôle de ces indices dans la mise en œuvre de la relation de discours est ensuite déterminé au moyen d'analyses sémasiologiques qui peut soit les valider comme des marqueurs, soit déterminer des faisceaux d'indices.

Liste d'axiomes pour inférer *Élaboration*

Axiome 1 Phase préparatoire, L93, page 79

$$\langle \tau, \alpha, \beta \rangle \wedge \text{prep}(\text{me}(\beta), \text{me}(\alpha)) > \text{Élaboration}(\alpha, \beta)$$

Axiome 2 Subtype, AL03, page 80

$$(?(\alpha, \beta, \lambda) \wedge \text{Top}(\sigma, \alpha) \wedge \text{Subtype}_D(\sigma, \beta, \alpha) \wedge \text{Aspect}(\alpha, \beta) > \text{Élaboration}(\alpha, \beta, \lambda)$$

Axiome 3 Subevent, VC10, page 87

$$(?(\alpha, \beta, \lambda) \wedge \text{Top}(\sigma, \alpha) \wedge \text{Subevent}_D(\sigma, \beta, \alpha) \wedge \text{Aspect}(\alpha, \beta) > \text{Élaboration}(\alpha, \beta, \lambda)$$

Axiome 4 Post Head Sentential, AL03, page 87

$$?(\alpha, \beta, \lambda) \wedge [\text{that}(x, p)](\alpha) \wedge p = \beta \rightarrow \text{Élaboration}(\alpha, \beta, \lambda)$$

Axiome 5 Notamment, VC10, page 154

$$(?(\alpha, \beta, \lambda) \wedge [\text{Notamment, }](\beta)) \rightarrow \text{Élaboration}(\alpha, \beta)$$

Axiome 6 En particulier, VC10, page 154

$$(?(\alpha, \beta, \lambda) \wedge [\text{En particulier, }](\beta)) \rightarrow \text{Élaboration}(\alpha, \beta)$$

Axiome 7 Plus précisément, VC10, page 165

$$(?(\alpha, \beta, \lambda) \wedge [\text{Plus précisément, }](\beta)) > \text{Élaboration}(\alpha, \beta)$$

Axiome 8 Structure énumérative, VC10, page 190

$$(?(\alpha, \beta, \lambda) \wedge \text{Énumération}(\alpha, \beta, \lambda) \wedge \text{Énumérathème}_{\text{événement}}[\alpha] > \text{Élaboration}(\alpha, \gamma, \lambda)$$

Axiome 9 Moyen, VC10, page 221

$$(?(\alpha, \beta, \lambda) \wedge \text{Top}(\sigma, \alpha) \wedge \text{Éventualité}_{\text{en cours de déroulement}}[\alpha] \wedge \text{Moyen}_D(\sigma, \beta, \alpha) > \text{Élaboration}(\alpha, \beta, \lambda)$$

L93 (Lascarides et Asher, 1993)

AL03 (Asher et Lascarides, 2003)

VC10 Nouveaux axiomes proposés

Index

- adverbe conjonctif, 131
- adverbe paradigmatissant, 130, 138, 260, 273
- annotation, 235, 236, 239, 301, 311
- approche ascendante, 9, 10, 30, 121, 235, 238, 240, 275
- approche descendante, 9, 16, 121, 235, 239, 275
- approche onomasiologique, 112, 123, 233, 285, 299
- approche sémasiologique, 112, 123, 233, 299
- Arrière-plan, 26, 43, 49, 89–91, 176, 202, 218, 224, 225, 245, 247, 249, 250, 257, 259, 271, 272
- attachement, 19, 22, 243, 252
- Attribution, 144, 245
- But, 243
- cadre de discours, 87, 146, 190, 242, 246, 271, 300
- Circonstance accompagnante, 215, 225, 226
- cohérence, 6–8, 18, 39, 74, 81, 83, 89, 106, 110, 112, 153, 243, 304
- cohésion, 5, 8, 15, 20, 46, 65, 69, 83, 114
- cohésion lexicale, 105, 112, 116, 123, 303
- connecteur, 103, 107, 111
- connecteur producteur, 111
- connecteur spécificateur, 111
- constituant, 18, 72, 73
- constituant complexe, 18
- constituant implicite, 25
- Conséquence, 200, 204, 212
- contenu sémantique, 6, 11, 13, 15
- contexte, 6, 8, 16, 20, 112, 121, 122, 162, 193, 211, 221
- Continuation, 73, 82, 143, 145, 146, 150, 154, 162, 165, 189, 226, 244, 293, 294
- Contraste, 27, 186, 189, 244, 312
- corpus, 127, 233, 235
- corrélats anaphoriques, 194, 195, 200
- d’abord, 171, 176, 179, 269
- dans un premier temps, 172, 177, 179, 270
- DRS, 15, 20
- DRT, 15
- effets sémantiques, 22, 23
- Élaboration, 13, 19, 30, 42, 46, 51, 57, 66, 68, 72, 79, 80, 85, 96, 116, 154, 165, 176, 190, 221, 246, 256
- Élaboration*, 165
- Élaboration d’entité, 89, 91, 92, 184, 246, 256, 261, 289
- en particulier, 133, 136, 147, 150, 152, 165, 166, 264, 265
- Encadrement, 245, 246
- Énumération, 187–189, 191–193, 201, 205
- état, 11, 16, 17
- éventualité, 16
- événement, 11, 17
- exemplification, 42, 47, 53, 62, 65, 67, 68, 246, 300
- Explication, 26, 43, 48, 49, 60, 142, 144, 176, 183–185, 190, 243, 291, 295, 296
- faisceau d’indices, 111, 112, 189, 193, 206, 275, 287, 303, 312
- Flashback, 176, 178, 245
- frontière droite, 19, 170, 193, 195
- gérondif, 208, 210, 217, 229, 272
- identification automatique, 300, 303, 304, 313
- indice, 10, 14, 20, 22, 100, 101, 111–113, 122, 150, 166, 177, 179, 205, 229, 239, 273, 301, 303, 309

- inférence, 7, 14, 23, 31, 57, 100, 108, 123, 183, 229
- introduceur de nouveau topique, 88, 146, 151, 191, 242, 246
- linguistique de corpus, 233
- Localisation temporelle, 245
- marqueur d'item, 188–190
- marqueur d'ouverture, 170, 180, 181, 269
- marqueur du discours, 100, 104, 111, 122
- marqueurs du discours, 101
- marqueurs potentiels, 111
- moyen, 211, 212, 218, 221, 222
- Narration, 24, 145, 146, 189, 202, 204, 218, 245
- notamment, 133, 136, 138, 150, 152, 165, 166, 260
- Parallèle, 27, 63, 157, 158, 195, 200, 244
- particulièrement, 131, 133, 136, 146, 149, 166
- plus particulièrement, 152, 154, 165, 267
- plus précisément, 152, 162, 165, 266
- précisément, 136, 148, 149, 166
- prédicat₁, 16
- prédicat₂, 23
- premièrement, 171, 174, 179
- propriété de distributivité, 81
- propriété de transitivité, 81
- réalisation linguistique, 113
- reformulation, 42, 67, 68, 80, 94, 96, 246, 286, 300
- règles de déclenchement, 20, 22, 23
- relation coordonnante, 13, 19
- relation de cohérence, 41
- relation de contenu, 13
- relation de discours, 6, 9–12, 14, 18, 20, 23
- relation structurelle, 12, 13, 27
- relation subordonnante, 13, 19
- relation véridique, 23
- représentation, 9
- représentation mentale, 6
- Résultat, 25, 48, 95, 176, 178, 218, 220, 222, 229, 243
- RST, 9, 11–14
- SDRS, 18, 20
- SDRT, 17
- segment de discours, 11, 18, 240
- segmentation, 105, 240
- signalisation, 14, 99, 112, 113, 116
- Source, 144
- structure du discours, 17
- structure énumérative, 181
- structure énumérative, 275
- subevent, 73, 86, 96
- subtype, 80, 85, 96
- Topique, 81
- topique, 9, 19, 25, 74, 88, 192
- traitement automatique des langues, 28, 113, 121, 303
- unité de discours élémentaire, 11
- voisins distributionnels, 307, 309, 312

Bibliographie

- ADAM, C. et VERGEZ-COURET, M. (2010). Signalling elaboration : Combining gerund clauses with lexical cues. *In Proceedings of the 8th MAD (Multidisciplinary Approaches of Discourse) : Multidisciplinary Perspectives on Signalling Text Organisation*, pages 80–90.
- ADAM, J.-M. et REVAZ, F. (1989). Aspects de la structuration du texte descriptif : les marqueurs d'énumération et de reformulation. *Langue française*, 81:59–98.
- AMOURETTE, C. (2006). L'expression du temps et de l'aspect par le participe présent et le gérondif. *In Les formes non-finies du verbe. Tome 1*, volume 19, pages 147–166. Travaux linguistiques du CERLICO.
- ASHER, N. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht, Kluwer.
- ASHER, N., AURNAGUE, M., BRAS, M., SABLAYROLLES, P. et VIEU, L. (1995). De l'espace-temps dans l'analyse du discours. *Sémiotiques*, 9(Numéro spécial Théories sémantiques et modélisation):11–62.
- ASHER, N., HARDT, D. et BUSQUETS, J. (2001). Discourse parallelism, ellipsis and ambiguity. *Journal of Semantics*, 18(1):1–25.
- ASHER, N. et LASCARIDES, A. (2003). *Logics of Conversation*. Cambridge University Press.
- ASHER, N., PRÉVOT, L. et VIEU, L. (2008). Setting the background in discourse. *Revue Discours*, 1:online.
- ASHER, N. et VIEU, L. (1995). Toward a geometry of common sense : a semantics and a complete axiomatisation of mereotopology. *In Proceedings of IJCAI95*, pages 846–852, San Mateo. CA : Morgan Kaufmann.
- ASHER, N. et VIEU, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, 115(4):591–610.
- ATALLAH, C. (2010). Les relations causales dans le discours. Mémoire de D.E.A., Université de Toulouse 2 Le Mirail.
- AURNAGUE, M. et VIEU, L. (1993). A three-level approach to the semantics of space. *In* ZELINSKI-WIBBELT, C., éditeur : *Semantics of Prepositions : From Mental Processing to*

- Natural Language Processing*, volume 3 de *Natural Language Processing*, pages 123–158. Berlin : Mouton de Gruyter.
- BALDRIDGE, J. et LASCARIDES, A. (2005). Probabilistic head-driven parsing for discourse structure. *In Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 96–103, Ann Arbor, Michigan. Association for Computational Linguistics.
- BEHRENS, B. et FABRICIUS-HANSEN, C. (2002). Connectives in contrast : a discourse semantic study of Elaboration on corpus research. *In HASSELGARD, H., JOHANSSON, S., BEHRENS, B. et FABRICIUS-HANSEN, C., éditeurs : Information Structure in a Cross-Linguistic Perspective*, pages 45–61. «Amsterdam», «New York» : Rodopi.
- BEHRENS, B. et FABRICIUS-HANSEN, C. (2005). The relation Accompanying Circumstance across languages. Conflict between linguistic expression and discourse subordination ? *In Proceedings from Third Contrastive Semantics/Pragmatics Conference*, International Studies University, Shangaï.
- BLUTNER, R. et ZEEVAT, H., éditeurs (2003). *Optimality Theory and Pragmatics*. Palgrave Studies in Pragmatics, Language and Cognition. Palgrave Macmillan, Londres.
- BORILLO, A. (1996). Exploration automatisée de textes de spécialité : repérage et identification de la relation lexicale d'hyperonymie. *LINX*, 34-35:113–124.
- BOUAYAD-AGHA, N. et SCOTT, D. R. (1998). Integrating constraints of layout and style for marking discourse structure. Unpublished paper.
- BOURIGAULT, D. (2002). Upéry : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. *In Actes de la Conférence Traitement Automatique du Langage Naturel*, pages 75–84, Nancy.
- BRAS, M. (2007). French adverb *d'abord* and discourse structure. *In AURNAGUE, M., KORTA, K. et LARRAZABAL, J., éditeurs : Language, Representation and Reasoning. Memorial Volume to Isabel Gómez-Trurruka*, pages 77–102. Bilbao : University of Basque Country Press.
- BRAS, M. (2008). *Entre relations temporelles et relations de discours*. Mémoire d'Habilitation à Diriger les Recherches, Université de Toulouse.
- BRAS, M. et ASHER, N. (1994). Le raisonnement non monotone dans la construction de la structure temporelle de textes en français. *In Actes du 9ème congrès de l'AFCEP RF-IA*, volume 2, pages 223–234, Paris.
- BRAS, M. et LE DRAOULEC, A. (2009). *D'abord* marqueur de structuration du discours. *Journal of French Language Studies*, 19(2):229–248.
- BRAS, M., LE DRAOULEC, A. et ASHER, N. (2009). A formal analysis of the french temporal connective *alors*. *OSLa*, 1(1):online.

- BRAS, M., LE DRAOULEC, A. et VIEU, L. (2001). French adverbial *puis* between temporal structure and discourse structure. In BRAS, M. et VIEU, L., éditeurs : *Semantic and Pragmatic Issues in Discourse and Dialogue : Experimenting with Current Theories*, volume 9 de *CRiSPi*, pages 109–146. Elsevier, Amsterdam.
- BRAS, M., PRÉVOT, L. et VERGEZ-COURET, M. (2008). Quelles relations de discours pour les structures énumératives ? In DURAND, J., HABERT, B. et LAKS, B., éditeurs : *Actes du Premier Congrès Mondial de Linguistique Française*, pages 1945–1964, Paris.
- BRAS, M. et SCHNEDECKER, C. (2009). Dans un (premier+second+nième) temps et les relations de discours : de l'élaboration à la contre-argumentation. In *Linguistic and Psycholinguistic Approaches to Text Structuring*, Paris.
- BUSQUETS, J. (2007). Discourse contrast : Types and tokens. In AURNAGUE, M., KORTA, K. et LARRAZABAL, J. M., éditeurs : *Language, Representation and Reasoning. Memorial Volume to Isabel Gómez-Txurruka*, pages 103–123. Bilbao : University of Basque Country Press.
- BUSQUETS, J., VIEU, L. et ASHER, N. (2001). La SDRT : une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum (Cohérence et relations de discours à l'écrit)*, 23(1):73–101.
- CARLSON, L. et MARCU, D. (2001). Discourse tagging reference manual. Rapport technique. www.isi.edu/marcu/discourse/tagging-ref-manual.pdf.
- CAWSEY, A. (1990). Generating communicative discourse. In DALE, R., MELLISH, C. et ZOCK, M., éditeurs : *Current Research in Natural Language Generation*, pages 75–102. Boston : Academic Press.
- CHAROLLES, M. (1983). Coherence as a principle in the interpretation of discourse. *Text*, 3:71–97.
- CHAROLLES, M. (1997). L'encadrement du discours - univers, champs, domaines et espace. *Cahier de Recherche Linguistique*, 6:1–60.
- CHAROLLES, M., LE DRAOULEC, A., PÉRY-WOODLEY, M.-P. et SARDA, L. (2005). Temporal and spatial dimensions of discourse organisation. *Journal of French Language Studies*, 15(2):115–130.
- CORNISH, F. (1990). Anaphore pragmatique, référence, et modèles du discours. In G. KLEIBER et J. E. TYVAERT, éditeurs : *L'anaphore et ses domaines*, pages 81–96. Paris, Klincksieck.
- CORNISH, F. (2003). The roles of (written) text and anaphor-type distribution in the construction of discourse. *Text*, 23(1):1–26.
- CRÉVENAT-WERNER, D. (2003). Quand *notamment* engage la partie. In COMBETTES, B., SCHNEDECKER, C. et THEISSEN, A., éditeurs : *Ordre et distinction dans la langue et le discours, Actes du Colloque International de Metz (18, 19, 20 mars 1999)*, pages 169–183. Paris, Champion.

- DALE, R. (1991). Exploring the role of punctuation in the signalling of discourse structure. *In Workshop on Text Representation and Domain Modelling : Ideas from Linguistics and AI*, pages 110–120, University of Berlin.
- DANLOS, L. (2004). Coréférence événementielle entre deux phrases. *In* LAPORTE, E., LECLÈRE, C., M. PIOT et SILBERZTEIN, M., éditeurs : *Syntaxe, Lexique et Lexique-Grammaire, Volume dédié à Maurice Gross, Linguisticae Investigationes Supplementa 24*. Amsterdam : John Benjamins.
- DANLOS, L. (2005). Partition of an entity with aspectuo-temporal operators. *In Proceedings of the Third International Workshop on Generative Approaches to the Lexicon*, Genève.
- DANLOS, L. (2008). Strong generative capacity of RST, SDRT and discourse dependency DAGs. *In* BENZ, A. et KÜHNLEIN, P., éditeurs : *Proceedings of Constraints in Discourse*, pages 69–95. Amsterdam : John Benjamins.
- DANLOS, L. et GAIFFE, B. (2000). Coréférence événementielle et relations de discours. *In Actes de la Conférence Traitement Automatique du Langage Naturel*, Lausanne.
- DANLOS, L. et HANKACH, P. (2008). Right frontier constraint for discourses in non canonical order. *In Proceedings of Constraints in Discourse*, pages 27–36, Potsdam.
- DAVIDSON, D. (1967). The logical form of action sentences. *In The Logic of Decision and Action*, pages 81–95. Pittsburg University Press, Pittsburg.
- DE MULDER, W. (2006). *Maintenant* : un connecteur token-réflexif? *In* MOLINE, E., STOSIC, D. et VETTERS, C., éditeurs : *Les connecteurs temporels du français, Cahiers Chronos*, numéro 15, pages 21–38. Rodopi.
- DEGAND, L. et SANDERS, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, 15(7-8):739–757.
- DEGAND, L. et SIMON, A. C. (2009). On identifying basic discourse units in speech : theoretical and empirical issues. *Revue Discours*, 4:online.
- DELIN, J., SCOTT, D. et HARTLEY, A. (1996). Language specific mappings from semantics to syntax. *In Proceedings of the International Conference on Computational Linguistics*, pages 292–297, Copenhagen.
- DI EUGENIO, B., MOORE, J. D. et PAOLUCCI, M. (1997). Learning features that predict cue usage. *In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 80–87, Morristown, NJ. Association for Computational Linguistics.
- DUCROT, O. (1983). *Le sens commun. Le dire et le dit*. Paris : Les Éditions de Minuit.
- EGG, M. et REDEKER, G. (2008). Underspecified discourse representation. *In* BENZ, A. et KÜHNLEIN, P., éditeurs : *Proceedings of Constraints in Discourse*, pages 117–138. Amsterdam : John Benjamins.

- ELHADAD, M. et MCKEOWN, K. R. (1990). Generating connectives. *In Proceedings of the International Conference on Computational Linguistics*, volume 3, pages 97–101, Helsinki.
- FABRICIUS-HANSEN, C. et BEHRENS, B. (2001). Elaboration and related discourse relations viewed from an interlingual perspective. Rapport technique, SPRIK report of the project Languages in Contrast, University of Oslo. <http://www.hf.uio.no/german/sprik>.
- FAYOL, M. (1997). On acquiring and using punctuation : A study of written french. *In COSTERMANS, J. et FAYOL, M., éditeurs : Processing Interclausal Relationships : Studies in the Production and Comprehension of Text*, pages 157–178. Lawrence Erlbaum Associates : Mahwah, New Jersey.
- FRASER, B. et MALAMID-MAKOWSKI, M. (1996). English and Spanish contrastive discourse markers. *In JASZCZOLT, K. et TURNER, K., éditeurs : Contrastive Semantics and Pragmatics : Discourse Strategies*, volume Volume II, pages 863–882. Pergamon Press.
- FREGE, G. (1892). Über sinn und bedeutung. *In Écrits logiques et philosophiques (1971)*, pages 102–126. Paris : Le Seuil.
- GETTRUP, H. (1977). Le gérondif, le participe présent et la notion de repère temporel. *Revue Romane*, 12(2):210–271.
- GIVÓN, T. (1983). *Topic Continuity in Discourse. A Quantitative Cross-Language Study*. Amsterdam / Philadelphia : Benjamins.
- GOLDMAN, A. I. (1970). *A Theory of Human Action*. Prentice Hall, Englewood Cliffs, NJ.
- GÓMEZ-TXURRUKA, I. (2003). The natural language conjunction *and*. *Linguistics and Philosophy*, 26(3):255–285.
- GOSSELIN, L. (1996). *Sémantique de la temporalité du français*. Louvain La Neuve, Duculot.
- GREFENSTETTE, G. (1994). Corpus-derived first, second and third-order word affinities. *In Proceedings of Euralex*, pages 279–290.
- GRICE, H. (1975). Logic and conversation. *In COLE, P. et MORGAN, J. L., éditeurs : Syntax and Semantics, v. 3 : Speech Acts*, pages 41–58. New York, Academic Press.
- GRIMES, J. (1975). *The Thread of Discourse*. The Hague : Mouton.
- GROSZ, B. J. et SIDNER, C. L. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- GROTE, B., LENKE, N. et STEDE, M. (1995). Ma(r)king concessions in English and German. *In Proceedings of the 5th European Workshop on Natural Language Generation*, pages 11–32, Leiden. Leiden University.

- GUIMIER, C. (1996). *Les adverbes du français : le cas des adverbes en -ment*. Collection l'essentiel français. Ophrys.
- HALLIDAY, M. (1985). *An Introduction to Functional Grammar*. Baltimore : Edward Arnold Press.
- HALLIDAY, M. et MATHIESSEN, C. (2004). *An Introduction To Functional Grammar*. Hodder Arnold.
- HALLIDAY, M. A. K. (1994). *An Introduction to Functional Grammar*. Arnold, Londres.
- HALLIDAY, M. A. K. et HASAN, R. (1976). *Cohesion in English*. Longman : London.
- HALMØY, O. (1982). *Le gérondif. Éléments pour une description syntaxique et sémantique*. Thèse de doctorat, University of Trondheim.
- HALMØY, O. (2003). *Le gérondif en français*. Collection l'Essentiel Français. Ophrys.
- HANSEN, M.-B. M. (1998). The semantic status of discourse markers. *Lingua*, 104:235–260.
- HARABAGIU, S. et MAIORANO, S. (1999). Knowledge-lean coreference resolution and its relations to textual cohesion and coherence. In *Proceedings of the Association for Computational Linguistics Workshop on Discourse/Dialogue Structure and Reference*, pages 29–38, University of Maryland.
- HARABAGIU, S. M. (1998). Wordnet-based inference of textual cohesion and coherence. In COOK, D. J., éditeur : *Proceedings of the International Florida Artificial Intelligence Research Society Conference*, AAAI Press, pages 265–269, Sanibel Island, Florida.
- HARABAGIU, S. M. (1999). From lexical cohesion to textual coherence : A data driven perspective. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 13(2):247–265.
- HARRIS, Z. (1968). *Mathematical Structures of Language*. New-York : JohnWiley & Sons.
- HIRST, G. (1981). Discourse-oriented anaphora resolution : a review. *American Journal of Computational Linguistics*, 7(2):85–98.
- HO-DAC, L.-M. (2007). *La position initiale dans l'organisation du discours : une exploration en corpus*. Thèse de doctorat, Université de Toulouse le Mirail.
- HO-DAC, L.-M., FABRE, C., REBEYROLLE, J. et PÉRY-WOODLEY, M.-P. (2009). From cues to markers : a corpus-driven approach to complex discourse organisation markers. In *Linguistic and Psycholinguistic Approaches to Text Structuring*.
- HOBBS, J. (1990). *Literature and Cognition*, chapitre 5. The coherence and structure of discourse, pages 83–114. Numéro 21. CSLI Lecture Notes.
- HOBBS, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3(1):67–90.

- HOBBS, J. R. (1985). On the coherence and structure of discourse. Rapport technique Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University.
- HOVY, E. (1988a). Planning coherent multisentential text. In *Proceedings of the Association for Computational Linguistics Conference*, pages 163–169, Buffalo.
- HOVY, E. (1988b). Two types of planning in language generation. In *Proceedings of the 26th ACL Conference*, pages 170–176, Buffalo.
- HOVY, E. (1989). Notes on dialogue management and text planning in the LILOG project.
- HOVY, E. H. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63(1 - 2):341–385.
- HOVY, E. H. et MAIER, E. (1991). Parsimonious or profligate : how many and which discourse structure relations. Unpublished Manuscript (online) www.isi.edu/natural-language/people/hovy/papers/93discproc.pdf.
- JACKIEWICZ, A. (2002). Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes. In *Colloque International sur la Fouille de Textes*, pages 95–105, Hammamet.
- JASINSKAJA, E. (2010). Modelling discourse relations by topics and implicatures : The elaboration default. In BENZ, A. et SIDNER, C. L., éditeurs : *Proceedings of Constraints in Discourse*, pages 61–80. Amsterdam : John Benjamins.
- JAYEZ, J. et ROSSARI, C. (1998). Discourse relations versus discourse marker relations. In *Processing of Workshop on Discourse Relations and Discourse Markers (Association for Computational Linguistics)*, pages 72–78, Montréal.
- KAMP, H. (1981). A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, pages 277–322. Groenendijk, J. A. G. and Janssen, T. M. V. and Stokhof, M. B. J., Amsterdam : Mathematical Centre.
- KAMP, H. et REYLE, U. (1993). *From Discourse to Logic : Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Boston : Kluwer, Dodrecht.
- KAMP, H. et ROHRER, C. (1983). Tense in texts. In BAUERLE, R., SCHWARZE, C. et von STECHOW, A., éditeurs : *Meaning, Use and Interpretation of Language*, pages 250–269. Berlin : Mouton de Gruyter.
- KAMP, H. et ROSSDEUTSCHER, A. (1994). Remarks on lexical structure and DRS construction. *Theoretical Linguistics*, pages 97–164.
- KEHLER, A. (2002). *Coherence, Reference and the Theory of Grammar*. CSLI Publications.
- KEHLER, A. (2004). Discourse coherence. In HORN, L. et WARD, G., éditeurs : *Handbook of Pragmatics*. Blackwell Reference Online.

- KINTSCH, W. (1977). On comprehending stories. In JUST, M. et CARPENTER, P., éditeurs : *Cognitive Processes in Comprehension*, pages 33–62. Lawrence Erlbaum Associates, Hillsdale, NJ.
- KLEIBER, G. (2007). La question temporelle du gérondif : simultanéité ou non ? In LAMBERT, F., MOREAU, C. et ALBRESPIT, J., éditeurs : *Les formes non finies du verbe -2- Travaux linguistiques du CERLICO*, volume 20, pages 109–123. Presses Universitaires de Rennes, Rennes.
- KLEIBER, G. (2008). Le gérondif : de la phrase au texte. In BERTRAND, O., PRÉVOST, S., CHAROLLES, M., FRANÇOIS, J. et SCHNEDECKER, C., éditeurs : *Discours, diachronie, stylistique du français. Études en hommage à Bernard Combettes*, pages 107–123. Berne : Peter Lang.
- KLEIBER, G. et THEISSEN, A. (2006). Le gérondif comme marqueur de cohésion et de cohérence. In CALAS, F., éditeur : *Cohérence et discours*, pages 173–184. Presses de l'Université Paris-Sorbonne, Paris.
- KLEIBER, G. et VASSILIADIOU, H. (2009). Sur la relation d'élaboration : des approches intuitives aux approches formelles. *Journal of French Language Studies*, 19(2):183–205.
- KNOTT, A. (1996). *A Data-Driven Methodology for Motivate a Set of Coherence Relations*. Thèse de doctorat, University of Edinburgh.
- KNOTT, A., OBERLANDER, J., O'DONNELL, M. et MELLISH, C. (2001). Beyond elaboration : the interaction of relations and focus in coherent text. In SANDERS, T., SCHILPEROORD, J. et SPOOREN, W., éditeurs : *Text Representation : Linguistic and Psycholinguistic Aspects*, pages 181–196. Amsterdam : John Benjamins.
- KORTMANN, B. (1991). *Free Adjuncts and Absolutes in English*. New York : Routledge.
- LAMBRECHT, K. (1994). *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge University Press.
- LASCARIDES, A. et ASHER, N. (1993). Temporal interpretation, discourse, relations and commonsense entailment. *Linguistics and Philosophy*, 6(5):437–493.
- LE DRAOULEC, A. (1999). Subordonnées temporelles et présupposition : quand la pragmatique s'en mêle. *Revue de Sémantique et Pragmatique*, 5:91–105.
- LE DRAOULEC, A. et BRAS, M. (2006). Quelques candidats au statut de connecteur temporel. *Cahiers de Grammaire*, 30:219–237.
- LE DRAOULEC, A. et BRAS, M. (2007). *Alors* as a possible temporal connective in discourse. *Cahier Chronos*, 17:81–94.
- LIN, Z., KAN, M.-Y. et NG, H. T. (2009). Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 343–351, Singapour. Association for Computational Linguistics.

- LUC, C. (2000). *Représentation et composition des structures visuelles et rhétoriques du texte. Approche pour la génération de textes formatés*. Thèse de doctorat, Université Paul Sabatier, Toulouse 1, France.
- MAIER, E. A. et HOVY, E. H. (1993). Organizing discourse structure relations using metafunctions. In HORACEK, H. et ZOCK, M., éditeurs : *New Concepts in Natural Language Generation*, pages 69–86. Pinter Publishers, Londres.
- MANN, W. C. et THOMPSON, S. A. (1986). Relational propositions in discourse. *Discourse Processes*, 9:57–90.
- MANN, W. C. et THOMPSON, S. A. (1987). Rhetorical structure theory : a theory of text organisation. Rapport technique, Technical Report ISI/RS-87-190, Information Sciences Intitute.
- MANN, W. C. et THOMPSON, S. A. (1988). Rhetorical structure theory : toward a functional theory of text organization. *Text*, 8:243–281.
- MANN, W. C. et THOMPSON, S. A. (2001). Deux perspectives sur la théorie de la structure rhétorique. *Verbum (Cohérence et relations de discours à l'écrit)*, 23(1):9–29.
- MANN, WILLIAM C. AND TABOADA, MAITE (2007). *RST Web Site*.
- MARCU, D. (2000). The rhetorical parsing of unrestricted texts : a surface-based approach. *Computational Linguistics*, 26(3):395–448.
- MARTIN, J. R. (1992). *English Text. System and Structure*. Amsterdam : John Benjamins.
- MAYBURY, M. (1990). *Planning Multisentential English Text Using Communicative Acts*. Thèse de doctorat, Cambridge University.
- McKEOWN, K. R. (1985). *Text Generation : Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press.
- MOLINIER, C. et LÉVRIER, F. (2000). *Grammaire des adverbes*. Genève, Droz.
- MONTAGUE, R. (1974). *Formal Philosophy : Selected Papers of Richard Montague, Edited and with an Introduction by Richmond H. Thomason*. Yale University Press.
- MOORE, J. et SWARTOUT, W. (1990). Dialogue-based explanation. In PARIS, C., SWARTOUT, W. et MANN, W., éditeurs : *Natural Language in Artificial Intelligence and Computational Linguistics*, pages 3–48. Boston : Kluwer.
- MOORE, J. et WIEMER-HASTINGS, P. (2003). Discourse in computational linguistics and artificial intelligence. In GRAESSER, A., GERNSBACHER, M. et GOLDMAN, S., éditeurs : *Handbook of Discourse Processes*, pages 439–486. Erlbaum, Mahwah, NJ.
- MOORE, J. D. (1989). *A Reactive Approach to Explanation in Expert and Advice-Giving Systems*. Thèse de doctorat, University of California in Los Angeles.

- MOORE, J. D. et POLLACK, M. E. (1992). A problem for rst : The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.
- MORTUREUX, M.-F. (1993). Paradigmes désignationnels. *Semen*, 8:123–141.
- NØJGAARD, M. (1995). *Les adverbes français - Essai de description fonctionnelle*, volume Tome II. Munksgaard, Copenhagen.
- NOLKE, H. (1983). Les adverbes paradigmatisants : fonction et analyse. *Revue Romane*, Numéro Spécial 23:191p.
- OATES, S. L. (1999). State of the art report on discourse markers and relations. Rapport technique, ITRI-99-08.
- PACELLI-PEKBA, T. (2003). Connecteurs et relations de discours : les cas de *quand*, *encore* et *aussi*. *Cahiers de Linguistique Française*, 25:237–256.
- PARIS, C. (1990). Generation and explanation : Building an explanation facility for the explainable expert systems framework. In PARIS, C., SWARTOUT, W. et MANN, W., éditeurs : *Natural Language in Artificial Intelligence and Computational Linguistics*, pages 49–82. Boston : Kluwer.
- PÉROZ, P. (2001). La construction des valeurs sémantiques de *d’abord* en position de détachement gauche. *Buletin Ştiinţific «Fascicula Limbi Moderne»*, 15(A):59–72.
- PÉRY-WOODLEY, M.-P. (2000). Une pragmatique à fleur de texte : approche en corpus de l’organisation textuelle. Mémoire d’Habilitation à Diriger les Recherches, Université Toulouse-Le Mirail, Carnet de Grammaire 8.
- PÉRY-WOODLEY, M.-P. (2001). Présentation. *Verbum (Cohérence et relations de discours à l’écrit)*, 23(1):3–8.
- PÉRY-WOODLEY, M.-P., ASHER, N., ENJALBERT, P., BENAMARA, F., BRAS, M., FABRE, C., FERRARI, S., HO-DAC, L.-M., LE DRAOULEC, A., MATHET, Y., MULLER, P., PRÉVOT, L., REBEYROLLE, J., VERGEZ-COURET, M., VIEU, L. et WIDLÖCHER, A. (2009). ANNODIS : une approche outillée de l’annotation de structures discursives. In *Actes de la Conférence Traitement Automatique du Langage Naturel*, page Online, Senlis.
- PIÉRARD, S. et BESTGEN, Y. (2006). Validation d’une méthodologie pour l’étude de marqueurs de la segmentation dans un grand corpus de textes. *Revue TAL*, 47(2):89–110.
- POLANYI, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, 12:601–638.
- POLANYI, L., CULY, C., van den BERG, M., THIONE, G. L. et AHN, D. (2004). Sentential structure and discourse parsing. In *Proceedings of the Association for Computational Linguistics Workshop on Discourse Annotation*, pages 80–87, Barcelone.

- POLANYI, L. et SCHA, P. (1984). A syntactic approach to discourse semantics. In *Proceedings of the International Conference on Computational Linguistics*, pages 413–419, Stanford, California.
- PORHIEL, S. (2007). Les structures énumératives à deux temps. *Revue Romane*, 42(1):103–135.
- PRASAD, R., MILTSAKAKI, E., DINESH, N., LEE, A., JOSHI, A., ROBALDO, L. et WEBBER, B.-L. (2007). The Penn Discourse TreeBank 2.0. annotation manual. Rapport technique, IRCS, Institute for Research in Cognitive Science, University of Pennsylvania.
- PRASAD, R., MILTSAKAKI, E., DINESH, N., LEE, A., JOSHI, A. et WEBBER, B. (2006). The Penn Discourse TreeBank 1.0. annotation manual. Rapport technique, IRCS Technical Report IRCS-06-01, Institute for Research in Cognitive Science, University of Pennsylvania.
- PRÉVOT, L. (2004). *Structures sémantiques et pragmatiques pour la modélisation de la cohérence dans des dialogues finalisés*. Thèse de doctorat, Université Paul Sabatier, Toulouse.
- PRÉVOT, L., VIEU, L. et ASHER, N. (2009). Une formalisation plus précise pour une annotation moins confuse : la relation d'élaboration d'entité. *Journal of French Language Studies*, 19(2):207–228.
- QUIRK, R. et GREENBAUM, S. (1973). *A Concise Grammar of Contemporary English*. New York : Harcourt Brace Jovanovich Inc.
- REBOUL, A. et MOESCHLER, J. (1998). *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*. Paris, Armand Colin.
- REESE, B., DENIS, P., ASHER, N., BALDRIDGE, J. et HUNTER, J. (2007). Reference manual for the analysis and annotation of rhetorical structure. Rapport technique, University of Texas at Austin.
- REICHMAN, R. (1978). Conversational coherency. *Cognitive Science*, 2:283–327.
- RIEGEL, M., PELLAT, J.-C. et RIOUL, R. (1994). *Grammaire méthodique du français*. Quadrigé.
- ROOTH, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.
- ROSSARI, C. (2000). *Connecteurs et relations de discours : des liens entre cognition et signification*. Presses Universitaires de Nancy.
- ROULET, E., AUCLIN, A., MOESCHLER, J., RUBATTEL, C. et SCHELLING, M. (1985). *L'articulation du discours en français contemporain*. Lang : Berne.
- ROZE, C., DANLOS, L. et MULLER, P. (2010). LEXCONN : a French lexicon of discourse connectives. In *Proceedings of the 8th MAD Multidisciplinary Perspectives on Signalling Text Organisation*, pages 114–125, Moissac.

- SÆBØ, K. J. (2006). The structure of criterion predicates. In DÖLLING, HEYDE-ZYBATOW et SCHÄFER, éditeurs : *Event Structures in Linguistic Form and Interpretation*, pages 127–147. Berlin : Mouton de Gruyter.
- SANDERS, T. et NOORDMAN, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29:37–60.
- SANDERS, T. et SPOOREN, W. (2001). Text representation as an interface between language and its users. In SANDERS, T., J., S. et SPOOREN, W., éditeurs : *Text Representation - Linguistic and Psycholinguistics Aspects*, pages 1–26. Amsterdam : John Benjamins.
- SANDERS, T. J. M., SPOOREN, W. P. et NOORDMAN, L. G. M. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, 15:1–35.
- SCHIFFRIN, D. (1987). *Discourse Markers*. Cambridge University Press.
- SCHNEDECKER, C. (2001a). Adverbes ordinaux et introducteurs de cadre - aspects linguistiques et cognitifs. *Linguisticae Investigationes*, 24(2):257–287.
- SCHNEDECKER, C. (2001b). Les corrélats anaphoriques *l'un/l'autre, le premier/le second* : aspects cohésifs de la référence «en stéréo». In ANDERSONS, H. et NOLKE, H., éditeurs : *Macro-syntaxe et macro-sémantique, Actes du Colloque International d'Arhus*, pages 257–283. Berne : Peter Lang.
- SCHNEDECKER, C. (2004). Particularités référentielles et (macro)-syntaxiques des textes à double ou à «multi-topique» - description linguistique et perspectives cognitives. In PORHIEL, S. et KUNGLER, D., éditeurs : *L'unité texte, Perspectives*, pages 251–274. Pleyben.
- SCHNEDECKER, C. (2007). *De l'un à l'autre et réciproquement... Aspects sémantiques, discursifs et cognitifs des pronoms anaphoriques corrélés l'un/l'autre et le premier/le second*. De Boeck Duculot.
- SCOTT, D. et DE SOUZA, C. (1990). Getting the message across in rst-based text generation. In DALE, R., MELLISH, C. et ZOCK, M., éditeurs : *Current Research in Natural Language Generation*, pages 47–73. Academic Press, London.
- SHIPHERD, H. R. (1926). *The Fine Art of Writing*. The Macmillan Co., New York.
- SIEPMANN, D. (2005). *Discourse Markers Across Languages. A Contrastive Study of Second-Level Discourse Markers in Native and Non-Native Text with Implications for General and Pedagogic Lexicography*. London & New York : Routledge.
- SUBBA, R. et DI EUGENIO, B. (2009). An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 566–574, Boulder, Colorado. Association for Computational Linguistics.
- THOMPSON, G. (2004). *Introducing Functional Grammar*. Arnold, 2^e édition.

- TURCO, G. et COLTIER, D. (1988). Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire. *Pratiques*, 57:57–79.
- VASSILIADOU, H. (2004). *Les connecteurs c'est-à-dire (que) en français et δeladè en grec. Analyse syntaxique et sémantico-pragmatique*. Thèse de doctorat, Université Marc Bloch, Strasbourg.
- VERBERNE, S., BOVES, L., COPPEN, P.-A. et OOSTDIJK, N. (2006). Discourse-based answering of why-questions. *Revue TAL*, 47(2):21–41.
- VERGEZ-COURET, M. (2006). Repérage en corpus des relations de discours et de leur marquage : le cas particulier de la relation d'*Élaboration*. Mémoire de D.E.A., Université de Toulouse 2 Le Mirail.
- VERGEZ-COURET, M. (2009). Le rôle de l'adverbe *notamment* dans la mise en place des relations de discours. *Journal of French Language Studies*, 19(2):249–268.
- VERGEZ-COURET, M., PRÉVOT, L. et BRAS, M. (2008). Interleaved discourse, the case of two-step enumerative structure. In *Proceedings of Constraints in Discourse*, pages 85–93, Potsdam.
- VIEU, L. (2007). On blocking : The rhetorical aspects of content-level discourse relations and their semantics. In AURNAGUE, M., KORTA, K. et LARRAZABAL, J. M., éditeurs : *Language, Representation and Reasoning. Memorial Volume to Isabel Gómez-Txurruka*, pages 263–282. Bilbao : University of Basque Country Press.
- VIEU, L., BRAS, M., ASHER, N. et AURNAGUE, M. (2005). Locating adverbials in discourse. *Journal of French Language Studies*, 15(2):173–193.
- VIEU, L., BRAS, M. et PRÉVOT, L. (2010). On the compositionality of temporal locating adverbial modification. In PODOGALLA, S. et AMSILI, P., éditeurs : *Actes des Journées Sémantique et Modélisation*, Nancy. 34–36.
- WEBBER, B., KNOTT, A. et JOSHI, A. (2000). Multiple discourse connectives in a lexicalized grammar for discourse. In *Proceedings of IWCS-00*.
- WHITTAKER, S. (2004). Aspects syntaxiques, sémantiques et textuels du comparatif nu en tête de phrase. *Romanische Forschungen*, 116(4):445–467.
- WHITTAKER, S. (2007). Le comparatif nu d'adverbes en -ment. *Travaux de linguistique*, 54:17–27.
- WILLIAMS, W. (1893). *Composition Rhetoric*. Heath and Co., Boston.
- WILSON, D. et SPERBER, D. (1990). Forme linguistique et pertinence. *Cahiers de Linguistique Française*, 11:13–53.
- WOLF, F. et GIBSON, E. (2005). Representing discourse coherence : A corpus-based study. *Computational Linguistics*, 31(2):249–287.